

Maastricht, Sarajevo and Rostock

Next week's referendum in France on the European treaty on monetary union is bound to be a milestone; the best hope is that it will mark the beginning of a better and wider Europe.

Is the European dream about to crumble? That could be the outcome of the French referendum on the Maastricht Treaty, due on Sunday 20 September. If even the chief instigators of the European Communities (EC) are able to turn their backs on Maastricht, do not others have a licence for saying that they had always known that the treaty was flawed and thus an excuse for walking away from it also? Last weekend's smart betting was that even politicians of a generous disposition would make of a French denial of Maastricht an opportunity to look quizzically at what the EC has done so far, picking and choosing among the legislative legacy of Brussels. Some might settle for a few environmental regulations, others for a package of tariff deals considered to be nationally advantageous. That would not be so much Europe *à la carte*, but a *smorgasbord*.

But there are several reasons why the omelette cannot be as casually unscrambled. For one thing, the process of economic integration has already gone too far. Who, now, could name as many computer manufacturers, or manufacturers of heavy electrical equipment, as there are member governments of the EC (twelve)? Even the EC's Common Agriculture Policy, based on the unreal promise that farmers' prosperity would always keep up with that of those in jobs of greater value (or "value added"), has at least dissuaded European farmers from cutting each others' throats (although they have often been ill-disposed towards each others' livestock). Who would relish the jungle that would spring up if there were a dozen parallel and protected farm systems?

Obstacles

Those are the obvious kinds of obstacles to turning back the clock to the early 1950s. More important is the cultural change that three decades have brought about. People in science know as well as any what is entailed. Where except at Geneva or Hamburg is high-energy physics practised seriously — unless at Frascati? (Heavy-ion specialists may have to go to Darmstadt or Nancy instead.) Astronomers, a self-supportive lot, are more in each others' pockets now than ever. Molecular biologists regularly make pilgrimage to Heidelberg or Cambridge. Senior people may be irritated at the airport waiting in-

involved. They should canvass opinion among their students, when they would learn that Europe has made young people's world a bigger oyster. Is not that something to be grateful for?

But would not cultural integration of this kind have happened even if there had never been an EC? That is what Maastricht's opponents ask. But they are wrong. The idea that a person can live and work wherever in Europe he or she chooses is a breathtaking innovation, conducive to a sense of liberty. So is the idea that a student (within reason) can study where he or she wishes. With a few decades more, there is every hope that Europe could spawn an intellectual enterprise comparable with that which made the fifteenth century come alive. Is that prospect in hazard if the French say NO a week on Sunday?

Flaws

The question would be less troublesome if the Maastricht Treaty were not indeed flawed. This journal has protested already that there is nothing in the treaty that will make Europe more democratic. The British government has been proclaiming in the past few months that it will force on Brussels the doctrine of subsidiarity, meaning that the European Commission cannot decide matters that could just as well be left to national governments; but there is nothing in that to suggest that national governments will be any more diligent than in the past at canvassing the opinions of their electorates. The possibility that even the Maastricht Treaty may now be at risk is a sign of how badly they have done in this regard.

Europe's money problems of the past few months are another illustration of Maastricht's defects. The essence of the treaty is a scheme for moving towards monetary union: a single currency, properly managed by a central bank. But what happens then to the relatively less well-favoured and less prosperous regions of Europe? Do the young people move elsewhere, leaving their elders in even greater discomfort? Either to prevent that happening or as a kind of welfare for the relatively indigent, it has always been supposed that monetary union would be accompanied by 'equalization' funds of some kind. But when Mr Jacques Delors, the president of the European Com-



mission, asked for them at Lisbon at the end of June, he was told he would have to wait.

That is why Europe's strategy now should not be to save Maastricht in all its detail, but to find a way of persuading France to endorse the principle while offering that the treaty will have to be changed. And that is a very curious business, for there is at least one compelling reason why the present treaty cannot stand: Denmark has already turned it down, in a referendum last July. Denmark's major partners have responded either with studied indifference or with the assumption that such a small country would eventually have to toe the others' line (perhaps one of the things about which the Danes were protesting). European interests would have been better served if the effort to reach an accommodation with Denmark had been undertaken quickly, so that voters in the French referendum would have seen that Maastricht is not written in stone, but in a kind of india-rubber. It should not have been beyond the wit of paid diplomats to have presented amendments to the treaty as if they were additions.

Of the many other reasons why the Maastricht Treaty cannot suffice, the present problems of ex-Yugoslavia are the best example. It is true that the treaty on European monetary union negotiated last December was accompanied by declarations of good intentions about cooperation on external relations; much has indeed been done in just half a year. But there is no framework within which the EC could have bargained, or can bargain, helpfully with Yugoslavia and its component parts. Membership of the EC? Formerly, the benefits of the agricultural policy would have been attractive, but reforms may dilute the appeal. External defence? That would be more persuasive if all twelve members of the EC were fulsome in support of the organization called Western European Union. Prosperity of some more general kind? That carrot may now be less persuasive.

The worst fear for the European dream is that its promise of prosperity may prove illusory. On the face of things, a community of 360 million people required by EC law to act competitively (which single-market regulations will require from next January) could not fail to create a division of labour of which Adam Smith would be proud. But will it come out like that? If Europe chooses to protect itself against technical innovations unless it has thought of them itself, it could as well become a living relic of the 1960s, the decade of good intentions. Or if it sets out to demonstrate its global civility by social and environmental regulations whose cost is never calculated, it could find that the overhead cost of the European dream defeats the purpose of making Europe prosperous. That, of course, is where Rostock on the Baltic is relevant; there is nothing as potent as economic promise denied to turn chauvinism to violent paths. □

Women in science

Discrimination against women in science is wrong but so is a quota system.

CURRENT predictions about science in the next century hold that the enterprise will require increasing numbers of female and minority researchers to compensate for declining numbers of white males choosing careers in research. Thus there are practical as well as philosophical reasons for promoting the advancement of women in science.

To that end, the biology directorate of the US National Science Foundation (NSF), which spends approximately \$2 million a year to support scientific meetings, has decided not to pay for conferences unless women are included as speakers.

The intention of the policy is laudable. Some one-third of the PhDs awarded in the United States in the biological sciences go to women (in the 1970s the figure was one-fifth), suggesting that there ought to be an ample pool of qualified women from which to draw. Therefore, NSF biology officials conclude, a conference without women on the programme is *prima-facie* evidence of discrimination, even if it is entirely unintentional. If a man (or woman) submits a meeting proposal with too few women, he or she will be told to remedy the apparent discrimination and submit the proposal again.

Sadly, this kind of nondiscrimination policy is in the event discriminatory itself, indirectly permitting the conclusion that women in science are not really as good as men. It is a phenomenon that has been reported elsewhere, for instance by black men and women admitted to elite colleges and universities under affirmative action policies. There is no evidence that sex is related to success in scientific research, and no inherent justification for holding women out for special treatment as part of a formal policy carrying the bludgeon of budgets.

Doubts can also be raised about legislation that has been introduced in the US House of Representatives to establish a commission to analyse what problems, if any, women in particular face in a life in research. It is a legitimate subject of inquiry. There is no doubt that fewer women than men are among those at the top of their profession and that this situation should change with the changing numbers of women in the workplace.

But proposals to deal with certain family issues that affect women more than men (child care and leave to tend to sick children, for instance) are already on the political agenda. Progress would be useful for all working women, but such issues have little to do with matters such as long hours in the laboratory that are part of the present culture of science. Change should, therefore, come from within the community. □

German institutes, industry seek to ease genetic engineering laws

Munich. The Max Planck Society has joined academic and industrial research groups in pressing the German government to change its gene technology laws — the most restrictive and bureaucratic in Europe. Although the organizations argue that the laws put an unnecessary burden on scientists and delay important experiments that pose no threat to the public, they concede that the political climate is against them and that no changes are likely in the next couple of years.

The problem arose in the autumn of 1989 when the government of Hessen, one of Germany's 16 states, blocked approval of a plant built by Hoechst to produce genetically engineered human insulin. The local ruling in effect prevented any genetic engineering for production purposes in Hessen.

At the time, there was no federal legislation controlling genetic engineering. Within months, however, the government had drawn up and won approval for a law consistent with a 1990 directive on genetic engineering from the European Communities (EC). The legislation is seen as a concession to the environmental Green party in the run-up to the 1991 elections, and critics say that too little thought was given to the scientific principles involved.

Researchers now face repetitive forms and delays in carrying out experiments. They complain that the law fails to make distinctions among experimental work in four safety bands. Experiments in the lowest band, which account for 80 per cent of cases, by definition pose no hazard to humans or the environment. But the federal law requires them to be treated in the same manner as experiments classified as low- and medium-risk: they must be done in certified laboratories after completion of a 30-page document describing any new experimental procedure and questionnaires on each experiment.

The society, Germany's largest independent research organization, would like to eliminate notification and documentation for experiments defined as carrying no risk. At the moment, says Peter Hofschneider, director of the Max Planck Institute for Biochemistry in Martinsried near Munich, it is as if the government required the same safety standards for bicycles as for aeroplanes. The ZKBS, the official body responsible for implementing the law, is not happy sifting through volumes of redundant information, and it has published a list of well-studied organisms and vectors

that are exempt from notification.

Hans-Georg Heidrich, safety officer at the Martinsried institute, says that the endless forms contribute nothing to safety. His unit actually ignores those for the lowest safety level and instead keeps relevant and appropriate written records, including those on risk assessment. That practice is acceptable to Bavarian officials but not to the



Max Planck's Hofschneider, left, and Heidrich.

Berlin-based pressure group, Genethisches Netzwerk. A spokesman, Wolfgang Lühr, says that biologists should be fully accountable for their experiments at all levels.

The society also wants separate forms for low-risk and medium-risk experiments that would collect more appropriate information and would eliminate the three-month delay now required between submission and approval of applications for no-risk and low-risk experiments. These delays are often extended for several months by requests for additional information. Max Planck scientists also want to regain the right to ship genetically engineered material to other countries, an important aspect of international collaborations.

Industry has objected to the federal gene law since its inception in July 1990. Dieter Brauer of Hoechst blames some of its flaws on ambiguities in the 1990 EC directive on biotechnology, which is difficult to integrate into national laws. Like academic researchers, industry complains about the bureaucratic obstacles that put German companies at a competitive disadvantage. The last straw has been a regulation allowing research only on small volumes; in Hessen, 'small' is defined as up to 10 l. A company whose developmental research requires larger volumes must comply with new, costly and time-consuming requirements before it can scale up.

These problems were acknowledged at a public hearing last February. But lack of confidence that things will change is forcing industrial genetic engineering out of Germany. BASF has built its US\$200 million

research institute for molecular biology in Worcester, Massachusetts. Bayer has decided to expand its pharmaceutical and medical research in the United States rather than at home. The gene law tipped the balance of Boehringer Mannheim's decision to build a large research and production centre in Minneapolis, Minnesota. And Hoechst has moved development of GMC-SF, hirudin and Factor XIII to its subsidiaries in the United States, France and Japan.

Industry has the option of moving out, whereas national institutes do not. As a result, academic researchers are maintaining pressure on the government despite holding out little hope of change before the next election in 1994.

Nevertheless, all sides believe that some tinkering with the rules may lighten the burden. They hope that the federal government may be persuaded to instruct individual states to interpret the law less strictly.

A more lenient approach will not alter the inherent contradictions of a gene law that judges a safe experiment to be unsafe — but it may help to contain the rising tide of bureaucracy. Officials may even be prepared to make some improvements to the weighty forms they distribute; for example, the ZKBS is extending a forthcoming list of exempted organisms and vectors, asking state governments to allow their unreported use.

Alison Abbott

California's money woes hit universities

Washington. Sixty-two days after it had run out of money, California's state government passed a budget last week that imposes cuts amounting to \$10.7 billion on state agencies, including a \$224 million (12 per cent) reduction from last year's funding for the nine University of California (UC) campuses. Although UC officials hope that revenue from a proposed tuition fee increase of 24 per cent will avert wholesale cuts, some research programmes are expected to feel the knife as the state looks for new ways to save money. Nearly 2,000 university employers chose a voluntary early retirement programme offered last year, and university officials are concerned that there are few left to do likewise this year. And the cost of operating the university system has risen by 30 per cent since 1988 despite state funding levels that have been flat or worse.

Christopher Anderson

NSF will experiment with fixed budgets in an effort to simplify grants process

Washington. Starting next month, mathematicians applying for support from the US National Science Foundation (NSF) will be part of an experiment intended to simplify the process by which researchers are funded, reduce the administrative burdens on both the foundation and the scientific community and increase the number of researchers receiving grants. In particular, applicants will not submit a proposed budget with their application and, if successful, will receive grants of a pre-determined amount.

The changes address two major problems. The first is how NSF can process half again as many research proposals from individual investigators as it did a decade ago, with no increase in staff (see charts). This has created a heavier workload for both NSF employees and for those outside who must review each proposal. In addition, a declining success rate — two in three applications are rejected — means that more of those reviews will be for naught. The second

problem is how to maintain the vitality of US academic research. Despite having more money, NSF is saying 'no' more often than in the past in most disciplines. And it uniformly gives those who succeed less money than they have requested.

Part of the crisis is self-generated: NSF is inviting applications for a growing portfolio of programmes covering minorities, women and young investigators as well as for special initiatives on such subjects as supercomputing, advanced materials and the environment. These typically have an even lower rate of success than does the typical research grant — a programme that intends to make a dozen awards, for example, may easily receive 200 applications. Each one, of course, must be submitted by the investigator's home institution before being processed by NSF and reviewed by as many as half-a-dozen colleagues.

The challenge for NSF is to modify its system without weakening the peer-review process that lies at its core. Although Walter

Massey, the NSF director, has mused in public about the concept of a block grant — giving an institution a lump sum and letting it decide how to divide the money — no such drastic changes are being considered. At the same time, a system that can spend \$10,000 to make a \$30,000 award clearly stands in need of revision.

The experiment being launched next month by NSF's mathematics divisions attempts to address several aspects of the problem. The awards will be for either \$20,000 or \$30,000 a year depending on the quality of the research; investigators may receive \$10,000 more for a graduate or postdoctoral student. The size of the grant will not be affected by an investigator's salary or the overhead costs charged by his or her university. Each of those features marks a break with tradition, in which budgets are set after individual negotiations reflecting the special circumstances of each investigator.

NSF expects to save enough money

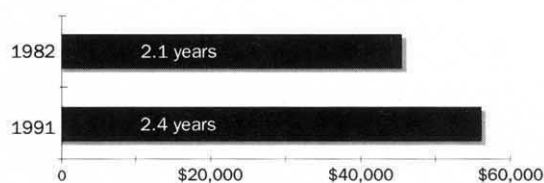
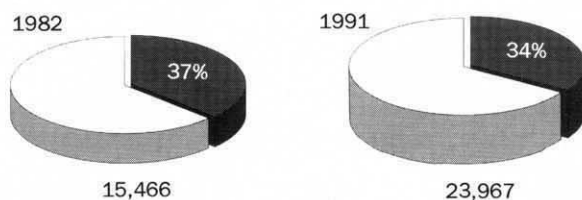
Longer odds, more competition . . . but a bigger payoff

NSF Research proposals*

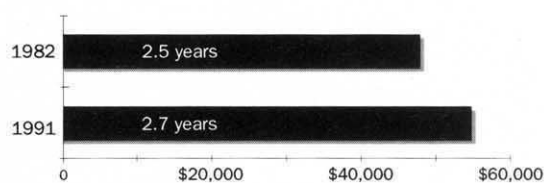
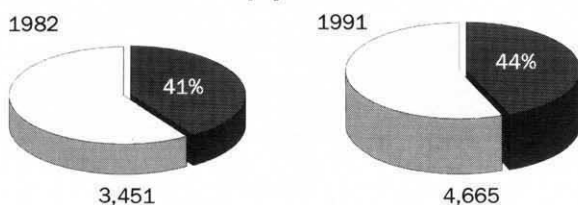
■ Funded □ Not funded

Average duration and size

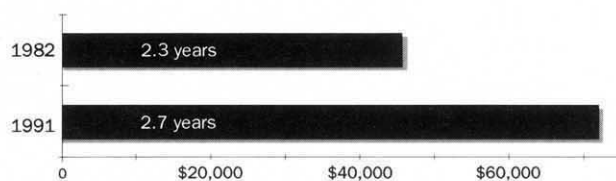
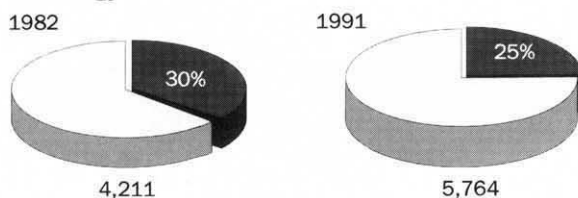
Overall



To mathematics and physical sciences



To biology



*Applications by individual investigators to research directorates

Canadian research council found guilty of job bias

under the new procedures to award about 20 per cent more grants. The beneficiaries will be the 120 or so mathematicians who in the past would have been turned down because of insufficient funds. The losers are the senior scientists whose grants will be trimmed to match one of the two flat rates. (An NSF grant historically pays two-ninths of a researcher's salary, on the assumption that the research is carried out during the summer term. The new grants are not intended to cover summer salary.)

"Fifteen years ago, I didn't worry about those people being turned down because they were not likely to make an important contribution", says Peter Bona, a mathematician at Pennsylvania State University and chairman of the mathematics advisory committee to NSF. "But the people being rejected today are a bit too good to ignore."

The chemistry division, according to its director, Kenneth Hancock, is considering a similar system with three tiers. The first would consist of standard grants, set at a flat rate to eliminate the need for negotiations. The second tier would provide small, starter grants to those new to the field, and the third would bestow large, long-term grants upon eminent senior researchers.

"It's a system in which a lot of people win", says Hancock. "We even found that we could fund at least as many grants as under the current scheme. The trick is that you have to sacrifice a little on the size of the average grant."

A flat rate will present investigators and their institutions with some hard choices, for example to sacrifice a higher salary for a graduate student or to spend money on travel rather than a new computer. Bona says that universities traditionally "make money" on mathematics departments because the scientists require little costly equipment or special facilities and that the new system may induce them to revise their bookkeeping and "to return some of that money" to mathematics.

The experiment, if successful, could be modified for other disciplines; in some ways, it extends efforts already under way to streamline the process. NSF officials have resolved to enforce an existing 15-page limit on grant applications, and the biology directorate has successfully used brief pre-proposals to screen applicants for multi-million-dollar training grants. The entire campaign, says Charles Brownstein, director of the office of planning and assessment, "is driven by a desire to avoid having highly qualified people become clerks".

The mathematics trial is scheduled to run for at least the 1993 fiscal year, which begins on 1 October, and possibly into 1994. Bona says the advisory committee will "be watching it like a hawk" and is prepared to make changes, or to jettison it, if there are signs that the new system is funding research of lower quality.

Jeffrey Mervis

Quebec. The National Research Council (NRC) of Canada, the country's largest research and development organization, has been found guilty of racial discrimination against an Indian-born physicist and ordered to review its human rights policy towards its employees. The ruling was made by the Canadian Human Rights Tribunal in



Chander Grover

a strongly worded decision issued on 21 August that recommended prosecution of several NRC employees for violating the country's Human Rights Act. NRC says that it is "surprised and shocked" by what is believed to be the first such investigation involving the council. Yet it seems unlikely to be the last. The complainant, Chander P. Grover, who arrived in Canada in 1978, claims that he was harassed by NRC even after the tribunal published its decision. Grover has made two further complaints that are being investigated by the Canadian Human Rights Commission, and has submitted two related grievances to the Public Service Staff Relations Board. A similar complaint — on the same grounds and against the same group of people at NRC — has also been submitted to the Human Rights Commission office by a former colleague of Grover's, Chinese physicist Tony Liu. A ruling is expected within the next few months.

The Canadian Human Rights Commission was established by the federal government in 1977 to investigate and resolve complaints of discrimination on such grounds as race, origin, religion, sex, age and disability. By last year it had more than 2,500 cases pending, 10 per cent claiming discrimination because of race and 12 per cent because of national origin.

Grover alleged that he was denied managerial, promotional and research opportunities in a series of incidents between September 1986 and August 1987. He works in the physics division at the NRC in Ottawa, and claimed discrimination on the basis of race, colour and national origin. In 1987 there were no visibly minority scientists within the senior management of the physics division.

After receiving degrees from the universities of Delhi and Paris VI and acquiring extensive research experience, Grover joined NRC in 1981 as an associate research professor within the physics division, and, ac-

cording to a previous director, "quickly became our leading expert in modern optics... his research work has brought him national as well as international recognition...".

In 1985 he became acting director of a new optics institute within the NRC that later became an independent National Optics Institute. He was appointed interim scientific director of the new institute, located in Quebec City, but in January 1987 turned down an offer to take the job on a permanent basis.

Grover's superiors then attacked his reputation, reduced his research activities, dismantled his research team, withheld his funding and left his future with the NRC uncertain, according to the tribunal. At one point, he was told he was being dismissed for disciplinary reasons, but the action was reversed the day before a staff hearing into his complaint.

In its 94-page decision, the tribunal found NRC's actions "flagrant and calculated to humiliate and demean the Complainant" and the cause of health problems for Grover and his family. NRC's own human rights adviser played a role in the debacle, the tribunal said: "After receiving in confidence Dr Grover's entire story about the complaints of his treatment...[she] then turns about and becomes the representative of the National Research Council throughout these complaints and the subsequent hearing". From the beginning of Grover's complaints to the last day of the hearing, the tribunal found, "NRC endeavoured to apply pressure on witnesses as well as control and prevent the introduction of some of the evidence".

NRC's treatment of Grover restricted his international activities and harmed his career. A last-minute cancellation of a trip to present a paper at a meeting in Boston of the Optical Society of America, to which he had recently been elected a fellow, "was professionally a great embarrassment for Dr Grover".

The tribunal called for a formal written apology to Grover from NRC's president within 15 days, to be published in an NRC publication, and an apology from the NRC president to the optical society. It ordered NRC to appoint Grover to at least a position of section head, to pay him lost wages, to cease its discrimination against him and to pay his legal fees and \$5,000 for emotional distress (the maximum allowed), including interest.

NRC is considering an appeal but has otherwise refrained from comment. It has 30 days to file an appeal, which would be heard by a division of Justice Department.

David Spurgeon

Japanese ministry requests large increase for competitive university grants

Tokyo. Japan's Ministry of Education, Science and Culture (MESC) has requested the largest increase in 15 years for its competitive research grants for university researchers. The proposed 16 per cent increase to ¥75 billion (US\$600 million) for the fiscal year that starts on 1 April 1993 follows a similarly large increase requested by the Ministry of International Trade and Industry (MITI) (see *Nature* 359, 4; 1992) and reflects a commitment by Japan's science-related ministries and agencies to achieving the government's goal of doubling its spending on research in the next few years.

The request comes a few weeks after the science council of MESC recommended that the competitive grants budget be increased by 50 per cent to ¥100 billion "at an early stage" (see *Nature* 358, 359; 1992). The request must be approved by the Ministry of Finance and the Diet, but its budgets generally undergo only minor trimming. Last year, for example, a request for 10.5 per cent more for competitive grants wound up as an increase of 9.7 per cent.

A rapidly rising competitive grants budget conforms to a standing policy by the ministry's bureaucrats to put a larger portion of the research grants budget on a competitive basis. Although the competitive grant budget has increased by

5–10 per cent a year in recent years, well above inflation, the budget for fixed maintenance grants of a few million yen (\$10,000–\$20,000) a year to all science and engineering professors has decreased in real terms almost every year since 1970

world's largest helical fusion device, under construction at the new national fusion research institute in Gifu Prefecture, would rise by more than 20 per cent, to ¥12.8 billion. The ministry has also asked for 25 per cent more for the Japan Society for the

Promotion of Science, which provides postdoctoral fellowships for Japanese and foreign researchers.

One programme being trimmed, by 9 per cent, is the TRISTAN electron-positron collider in the high-energy physics laboratory at Tsukuba science city, as attention shifts to the more powerful Large Electron-Positron (LEP) collider in Europe. The earthquake and volcanic eruption prediction programmes would receive a 5.5 per cent increase despite doubts about whether earthquake prediction is within the present capabilities of science (see

Nature 358; 361; 1992). The ministry's Antarctic research base gets a similar increase.

The large increases requested by MESC and MITI go some way towards reaching a goal approved by the cabinet earlier this year of doubling Japan's government research budget in the near future. But maintaining the present growth rates will require Japanese industry and the public to accept some form of extra taxes.

David Swinbanks

Highlights of budget proposal for MESC

	billion¥	% change from 1992
Grants-in-aid of research	75.0	+16.1
Government/industry research	15.3	+3.2
Donations from industry	50.1	+4.1
Nuclear fusion	12.8	+20.4
Accelerator physics (TRISTAN)	13.0	-9.0
Space science	20.8	-0.5
Astronomy (Hawaiian telescope)	4.0	+38.9
Earthquake & volcanic eruption prediction	2.4	+5.5
Antarctic research	3.7	+5.7
Japan Society for Promotion of Science	9.6	+24.8

(see *Nature* 339; 576; 1989).

But more money and greater competition are not the only changes needed in the grants system. Young university researchers say the process of reviewing competitive grant applications must be made more open; at present, small committees of senior academics review applications in secret and do not explain their actions.

MESC has also requested a large increase for an 8-metre infrared telescope being built in Hawaii. And spending on the

SKB announces UK redundancies and realignment

Munich and London. SmithKline Beecham has dismissed 150 British employees as part of a realignment of its research and development efforts. Although the redundancies within the 5,000-person work force were larger than expected, last week's announcement ends the uncertainties about the drug company's British operations since the merger between US-based SmithKline Beechman and UK-based Beecham in July 1989.

The company's eight research sites scattered across Britain will be combined into six as quickly as possible. The 2,000 researchers in the United States are not affected.

The long-expected reorganization leaves researchers in Britain focused mainly on the central nervous system, although the highly successful anti-infection programme built up by Beecham

will be maintained. All cardiopulmonary work, some of which was carried out in Britain, will now be done in the United States, and vaccine work will continue in Belgium.

But much of the company's work on peripheral systems, notably gastrointestinal, has been swept away by the changes. Twenty years ago, this area provided the mainstay of SmithKline's profits after Nobel prizewinner James Black developed for them a new class of drugs — the antiulcer H_2 receptor antagonists, the world's second most commonly prescribed drugs. Cimetidine, marketed by SmithKline as Tagamet, has had annual sales of more than US\$1 billion in the last six years.

But patent rights are running out, and a new class of more potent antiulcer drug — proton-pump inhibitors such as Astra's

omeprazole — is now on the market. SmithKline Beecham has decided not to continue its search for new therapeutic targets in this area.

The changes also mean an internal restructuring away from a project-orientated approach to research towards a departmental system. For example, groups previously assigned to a specific illness such as depression will become part of a large neuroscience department. SmithKline Beecham says that teams will be integrated into the new structure without senior scientists losing their status.

SmithKline Beecham is the fifth-highest corporate spender on research in Britain. Within the pharmaceutical industry, the £432 million it invested last year ranked second only to the £475 million spent by Glaxo.

Allison Abbott & Ian Mundell

Physicist running for president is accused of distorting science to fit guru's ideas

Washington. To those who want science to play a larger role in politics, John Hagelin is a reminder that such a development may be a mixed blessing.

Hagelin, a quantum physicist trained at Harvard University, is running for US president as the candidate of the Natural Law Party, whose motto is "bringing the light of science into politics". He and a supporting



John Hagelin

cast of local candidates will appear on the ballot in at least 35 states. (More than 300 people ran unsuccessfully under the party's banner in the British parliamentary election in June.)

He is by all accounts a gifted scientist, well-known and respected by his colleagues. He is a co-developer of one of the better-accepted unified field theories, known as the flipped SU(5) model. In May, he received an award for young innovators named after Jack Kilby, inventor of the integrated circuit. And his political platform is eminently sensible: practise only those social and economic policies that are supported by scientific data.

However, there is another side of Hagelin that disturbs many researchers. Hagelin is a

follower of Maharishi Mahesh Yogi, best known as the guru who taught the Beatles about transcendental meditation (TM), and is on sabbatical from Maharishi International University in Fairfield, Iowa, where students practise mass meditation as a way to ease many of the world's ills, from crime to stress. The home of the Natural Law Party is near the university and most of its members embrace the Maharishi's teachings.

Hagelin has been investigating a scientific mechanism to explain how TM can influence world events. The answer, he believes, lies in extending the grand unified theories of physics to human consciousness.

In the past several years, Hagelin has worked on integrating the SU(5) model, which does not include gravity, into the four-dimensional heterotic superstring model, which is currently considered one of the better prospects for a grand unified 'theory of everything'. Everything, in this case, may even include human consciousness. Two-page advertisements, with row after row of partial differential equations, appear regularly in US newspapers describing how the theoretical physics work of Hagelin and others explains the impact of TM on distant events. Hagelin often lectures on SU(5) and other unified field theories to both scientific and nonscientific audiences, mixed in with a lengthy discussion of TM.

Not surprisingly, the linkage of SU(5) with TM infuriates his former collaborators.

It is hard enough, they complain, to win scientific support for any type of unified theory. "A lot of people [Hagelin] has collaborated with in the past are very upset about this," says Jorge Lopez, a Texas A&M University physicist. "It's absolutely ludicrous to say that TM has anything to do with flipped SU(5)."

John Ellis, director of CERN's theoretical physics department, has asked Hagelin to stop mixing TM and SU(5). "I was worried about guilt by association," Ellis explains. "I was afraid that people might regard [Hagelin's assertions] as rather flaky, and that might rub off on the theory or on us."

Physicists are not the only scientists to take issue with Hagelin's mix of science and politics. One plank of his party platform calls on presidential candidates to undergo an electroencephalogram (EEG) brain scan that would purport to reveal their neurophysiological qualifications to hold office.

EEG scans, in use since the 1930s, "show the orderliness of the brain", he explains. "Science correlates that to intelligence, creativity, moral stability and broad comprehension." He says that he has had his own brain scanned (he claims an exceedingly orderly brain, in the top 1 per cent of those tested) and will release the results when his competitors do. EEGs, he says, "give us a look under a candidate's hood".

Jonathan Pincus, chairman of the neurology department at Georgetown University in Washington DC, says that researchers once hoped the results of EEG tests might somehow correlate to intellectual qualities. Although EEGs have remained an important tool for spotting neurological disorders, he says, "they have nothing whatsoever to say about a person's moral fibre".

Hagelin himself cites work by E. Roy John, director of the Brain Research Laboratory at the New York University Medical Center, to back his claims. But John says that Hagelin is "overselling" the technique. EEG brain scans have been shown to correlate to "a large number of subtle malfunctions", from senility to substance abuse, he says, "but qualities like moral stability and intelligence are simply not measured".

Even the Kilby award is a bit of a mystery. Few have heard of it, perhaps because it was created three years ago by the North Dallas Chamber of Commerce to draw attention to the area. Truman Cook, a chemical engineer who is a member of the award's executive committee, says that a member of the selection committee who practises TM proposed Hagelin for the award.

Ian Mundell

Christopher Anderson

Wellcome Trust to list jobs, grants on line

London. Job vacancies and funding sources in the biomedical sciences will be available electronically in the spring on a new database to be created by the Wellcome Trust.

The service, to be run from the trust's newly restored Centre for Medical Science in London, will be offered free via the JANET electronic mail system available to most British academic scientists. Although some informal advertising of vacancies already takes place on electronic bulletin boards, the Wellcome Trust's database is thought to be the first formal system on this scale in Britain.

Although the trust expects the jobs database to be self-supporting, the charges will be minimal to enable advertisers to keep positions on the system for longer. Users will be able to access vacancies according to subject, geographical description or any number of hierarchical measures.

The centre's director, Laurence Smaje, said that the charge for a job listing will

probably be between £10 (US\$20) and £20 per week; by comparison, a quarter-page in *Nature* costs £600 and twice that in *New Scientist*. Although the trust intends to promote the database vigorously, Smaje says that it is not intended to steal business from the weekly magazines and will not promote lectureships and industrial vacancies.

The first step will be to include all vacancies generated by Wellcome Trust or Medical Research Council awards, adding other information as available. While the trust will focus on Britain, organizations elsewhere may also submit material.

Although much of the information on grants is available from libraries, it is largely tied up in a mass of books or on internal databases such as that held by the Association of Medical Research Charities. As well as assembling this information, the grants database will provide such additional facts as the amount of money at a charity's disposal and policies on who may apply.

US university told to reveal unfunded NIH application

A state court in Washington has ruled that the University of Washington in Seattle must allow an animal welfare group to see a proposal by one of its faculty for an experiment on rhesus monkeys even though the idea was not funded and the experiment has not been done. In a decision on 17 August, a King County superior court judge ordered psychologist Gene Sackett to release a copy of the proposal to the Progressive Animal Welfare Society (PAWS), a local activist group. It is believed to be the first time a court in the United States has ordered the release of an unfunded grant application.

PAWS successfully argued that the Washington state public records laws permit the release of any grant proposal filed by a state-supported researcher regardless of the outcome. Had the proposal been funded by the National Institutes of Health (NIH), to which it had been sent, its contents could have been disclosed under the federal Freedom of Information Act, according to NIH officials.

Twenty-nine other organizations, including NIH and several other US universities, filed declarations on behalf of the University of Washington. They argued that release of the information violates an investigator's right to privacy and could reveal secret research formulas, designs or data, as well as setting a precedent for other state courts. The decision — if it stands — might also allow scientists to take ideas from a

colleague's application, refine them and then submit them for funding in competition with the original proposal, says Joanne Belk, NIH's acting freedom of information officer.

But the judge ruled that neither state law nor the First Amendment of the US constitution that protects freedom of speech prohibit the release of such unfunded applications. He did, however, allow the university to withhold budget figures, personal information about the investigator and some financial calculations. Although the state has unusually liberal disclosure laws, according to Laurel Wilkening, provost of the University of Washington, "these things tend to cascade" and could prompt similar orders to release research documents in other states.

The experiment, called "Effects of socialization on forebrain development", would have used 21 infant rhesus monkeys, 13 of which would have been deprived of contact with other monkeys shortly after birth. Sackett speculates that the isolation may cause self-destructive tendencies in the monkeys that would produce changes in their brains; such a finding might someday lead to treatments of children who exhibit similar behaviour. NIH declined to fund the proposal in June 1991 after widespread community and campus protest. University officials have not yet decided whether to appeal.

Christopher Anderson

Genetics-and-crime conference cancelled in clash with NIH

Washington. Organizers last week cancelled a controversial conference on genetics and crime after the US National Institutes of Health (NIH) refused to release money it had previously awarded for the meeting.

Attacked as racist by African-American groups, the conference was scheduled to bring together geneticists, ethicists and criminologists on 9-11 October to discuss whether there were genetic links to criminal behaviour. But critics complained that the promotional material seemed to assume such a controversial link and the NIH froze its \$78,000 grant (see *Nature* 358, 357; 1992) from the ethics, legal and social implications programme of the National Center for Human Genome Research. Although the conference's organizer, David Wasserman of the University of Maryland, convened a panel last month to address objections and submit an alternative list of speakers and topics, the debate only became more acrimonious.

Last week, NIH wrote a three-page letter to Wasserman accusing him of "irresponsible" behaviour. It called the conference "fatally flawed" and rejected the alternative agenda that Wasserman had proposed. Although the conference, sponsored by the university's Institute for Philosophy and Public Policy, can no longer be held on schedule, Wasserman still wants NIH to "recognize its obligation to release the money".

Christopher Anderson

NEWS IN BRIEF

Washington. US biotechnology companies are concerned that the US government is about to lump them in a class with Wall Street investment firms, forcing them to disclose internal details of their research programmes.

The Securities and Exchange Commission (SEC) has warned several cash-rich biotechnology companies without products on the market that it may label them as 'investment companies' because the companies have large cash reserves — often more than \$100 million — and tend to buy large amounts of high-yielding securities until the money is needed to meet research costs. Start-up companies may take years to exhaust such reserves.

This practice has apparently caught the notice of the SEC, which regards such investments as one sign that a company should be monitored more closely. Should the SEC follow its proposed redesignation, affected biotechnology companies would have to comply with yet another layer of regulation, possibly including higher taxes,

disclosure requirements and financial restrictions. The Industrial Biotechnology Association has met SEC officials on the issue and will discuss a response at a meeting of its board of directors on 23 September.

C.A.

Munich. The Collège de France, the elite Paris university whose select band of professors enjoy great prestige in France, is opening its doors to the world. Once the bastion of nationalist culture, the Collège is for the first time allowing foreigners to be considered for its 52 permanent chairs, 22 of which are occupied by scientists. This decision follows the success of a single chair, established in 1989, that offers a yearly contract to someone from another country in the European Communities. Professors are not required to have any specific qualifications, nor are they required to follow any particular curriculum. Their lectures, which must be changed every year, are open to the public.

A.A.



Barbara McClintock

Washington. Barbara McClintock, whose discovery of transposable genetic elements won her a Nobel Prize in 1983 and a crowning place in scientific history, died last week after a short illness. Her

pioneering work on 'jumping genes' in maize in the 1940s and 1950s predated all but the vaguest descriptions of genes themselves. Associated with the Cold Spring Harbor (NY) Laboratory for more than 50 years, she continued to work in genetics, inspiring generations of scientists. A collection of essays by leading geneticists on McClintock's scientific and personal influence was published earlier this year to mark her ninetieth birthday (see *Nature* 358, 631; 1992).

C.A.

Scope of Human Frontiers

SIR — As a member of the Council of Scientists of the Human Frontier Science Program (HFSP) since the programme was launched three years ago, I appreciate the recent discussion in *Nature* about an issue crucial to the programme and its future, namely its scientific scope. I have always strongly supported the original intention behind the programme, expressed by Dr Akiyoshi Wada (*Nature* 357, 356; 1992) and endorsed editorially (358, 525 & 527; 1992), that the broad interdisciplinarity of the HFSP is exactly what makes the programme a frontier science programme. Take that away and the programme is reduced to a traditional, albeit international, funding agency for the biological sciences. There are two important questions that seem to have been left hanging in the air.

First, although not at all surprising to insiders, it is a matter of real concern that Dr Edward Rall (present chairman of the Council of Scientists), Dr Joseph Varner (former member of the Council of Scientists) and Sir James Gowans (retiring Secretary-General of the Human Frontier Science Program Organization) apparently agree on an important point of immediate and potentially serious practical consequence. They apparently all think that Wada's article in *Nature* has created confusion among the applicants in the next round of applications for the HFSP for which the closing date is the end of September 1992. Their position amounts to (1) a warning to potential applicants that they should not submit proposals to the HFSP that fit Wada's description of the broad interdisciplinarity of the programme and (2) a flat denial of the broad interdisciplinarity of the programme itself.

I should like to emphasize that Wada has not created any confusion among potential applicants. His presentation of the scope of the programme agrees exactly with the published information on the programme, including all published information to potential applicants. Wada has merely, and in my view correctly, drawn attention to the fact that the HFSP is a truly interdisciplinary scientific programme whose scope includes the many disciplines that currently collaborate with the biological sciences to elucidate the functions and mechanics of living organisms, such as physics, chemistry, engineering, the cognitive sciences and the study of behaviour as caused by information-processing. I recommend therefore that potential applicants to the HFSP should not refrain from submitting truly interdisciplinary research proposals that fit the published description of the programme's scope.

My second point is simple and has in effect already been made. The issue over the broad interdisciplinarity of the Human Frontier Science Program is not an issue of "Japan against the rest of the world": it is an issue between those who wish to maintain the HFSP's hallmark as a true frontier science programme and those who do not.

Niels Ole Bernsen

*Centre of Cognitive Science,
Roskilde University,
PO Box 260,
DK-4000 Roskilde, Denmark*

Occupational risks

SIR — Attractive though they are, the technical properties of ultra-thin man-made fibres pointed out by Paul Calvert (*Nature* 357, 365; 1992) should not hide the potential — at least for those fibres resistant to biological degradation *in vivo* — for related occupational risks to workers.

Fortunately, most reinforcing fibres hitherto produced in quantity have, as Calvert pointed out, been of diameter 10 µm or more; the practical risk from occupational or other exposure to their airborne dusts remains doubtful. But work on fibres other than asbestos has shown the morphology and biological persistence of fibrous materials to be of greater significance in relation both to pneumoconiosis and, more seriously, to mesothelioma, than their chemical constitution.

A need for stringent precautions in preventing occupational exposure to the dusts of these thinner materials might well result in cost increases in manufacture that would outweigh the "dramatic reduction in production costs" hypothesized by Calvert.

Gerald V. Coles

*Occupational Hygiene Unit,
Deakin University,
Geelong, Victoria 3217, Australia*

Czech science

SIR — I was disgusted by Dr Z. Vaněk's letter about the recent problems in Czechoslovak science (*Nature* 357, 10; 1992). His allegations that the newly introduced granting system in the institutes of the Academy of Sciences is based on "family favours" made me really furious. There are many problems in our science, but the introduction of an internal granting system by the academy was clearly the most positive achievement of the past two years. This was the first attempt in our country to lay new

foundations for selective funding of good projects.

It is not the fault of the academy that it has not yet been possible to create a more general granting agency to fund research outside the academy. The academy's internal granting agency began to operate last year on the anonymous peer review principle; scientific and administrative bodies of the agency were democratically elected by the scientific community. About 40 per cent of all applicants were given grants. I would like to point out that there have been truly deep democratic reforms in the Czechoslovak Academy of Sciences since 1989; to my taste the present system is perhaps too democratic. I say this even though I am just an ordinary scientist, not a member of any academy or granting agency bodies. Clearly Vaněk did not get a grant for his project.

Similarly untrue is Vaněk's statement that existing institutes are being forced to split into smaller institutes in order to create more posts for directors. This has not happened, although many scientists believe that some large and heterogeneous institutes could perhaps be divided into smaller and more efficient independent units. I can assure you that at present few people would like to be a director. Vaněk's attacks on independent international boards that should evaluate the efficiency of the institutes is characteristic of those who are afraid of the results of such evaluation. Finally, I would like to express my disappointment that you failed to check the real state of affairs before publishing Vaněk's claims.

Václav Hořejší

*Institute of Molecular Genetics,
Czechoslovak Academy of Sciences,
Vítězská 1083,
142 20 Praha 4, Czechoslovakia*

Vaccine errors

SIR — There are several errors in your news story about the Sclavo Research Institute (*Nature* 357, 6; 1992). In the last paragraph, *Haemophilus influenzae* b is referred to as a virus, when it is in fact a bacterium. Second, what is described as "a vaccine against recombinant pertussis toxin" is actually a vaccine consisting of a recombinant pertussis toxin.

Furthermore, many would take issue with the "association with neurological damage" of the "current whooping cough vaccine". Lastly, the bacterium *Helicobacter pylori* is misspelled.

Janet R. Gilsdorf

Carl F. Marrs

*University of Michigan Medical Center,
C. S. Mott Children's Hospital,
1500 E. Medical Center Drive,
Ann Arbor, Michigan 48109-0244, USA*

Popper: success or failure?

SIR — Popper's own estimation of himself does not square with Bondi's¹. "Here I [Sir Karl Popper] am being showered with honours as no professional philosopher before me; yet three generations of professional philosophers know nothing about my work"².

Popper is a remarkable philosopher, but his success — and failure — is even more remarkable. Most human existence has been in closed societies. We are lucky: we live in an open society. Popper more than anybody this century has made us aware of its importance, and its relationship with science. He has also shown that both owe their existence to critical debate and to our collective preparedness, as citizens and scientists, to reason, debate, conjecture and criticize ideas.

But he has also failed. Popper notes that the Greeks invented the tradition of critical debate and the open society³. But from Popper, no tradition has arisen for their study. There is no school of Popper: the only popperians are scientists, like Bondi, not philosophers or those studying society.

Indeed, Popper's failure has been worse. There has been another equally remarkable story: the rise in much of contemporary philosophy and sociology of the conclusion that criticism and open debate are of little or no importance, either in society or science. Instead, scholars of science widely attribute its success, together with that of our open society, to such things as political power, ideologies, and 'social contingencies'. Does the failure of Popper's ideas among those studying science and society matter? I think it does. I focus upon its relevance to science.

(1) Popper and the contemporary study of science cannot both be right. One of them has to be outlandishly wrong. This disjunction is ignored but it should be faced. There is no necessary reason to respect all scholarship. If Popper is right, then the study of science has to be a study of criticism and debate. By implication, those talking about things such as 'social contingencies' must be doing something other than studying science. Perhaps, under the guise of 'theoretical speculation', they are reading things into science that belong more to their imaginations than science. As much as there is a pseudoscience of the stars, there can be a pseudoscience of science. Scientists should not be afraid of saying so.

(2) Does Popper's failure have wider, unnoticed implications for scientists? For Popper, science is one of the greatest achievements of our society — one of its unending quests. For those who see the

success of science in terms of social contingencies, it is just another social activity — an overprivileged one. Whichever view of science prevails will shape its long-term welfare. Science depends upon public funds. Its success therefore rests on the value politicians and bureaucrats put upon science. But the chances are that they were once as students taught the contemporary, not the popperian, view of science. If this is the case, Popper's failure may have cast a shadow over science and its unending quest.

John R. Skoyles

6 Denning Road,
London NW3 1SU, UK

1. Bondi, H. *Nature* **358**, 363 (1992).
2. Popper, K. R. in *In Pursuit of Truth* (ed. Levinson, P.) 249–289 (Humanities, New York, 1982).
3. Popper, K. R. *Conjectures and Refutations* (Routledge and Kegan Paul, London, 1963).

How secure?

SIR — You say (*Nature* **358**, 356; 1992) that plutonium could be stored in "internationally secure repositories". Unfortunately repositories could not be safer than the politicians, dictators or rebels who would look after them.

Richard Baker

22 Cromwell Road,
Kingston KT2 6RE, UK

Leprosy vaccine

SIR — K. S. Jayaraman¹ implies that BCG vaccine has a protective effect against leprosy in Venezuela and that this is "the first demonstrated effect of BCG on leprosy in the New World". But this conclusion is drawn from a retrospective case-control study, which is different from the randomized, double-blind trial comparing the effect of BCG alone with that BCG and *Mycobacterium leprae* and which did not include a placebo group².

The failure to include a placebo group in this and in the current trial in Malawi is all the more surprising given the contradictory results of the previous BCG trials in Uganda, New Guinea and Burma. Moreover, any useful vaccine should be effective against the borderline and lepromatous multibacillary form of the disease that is considered responsible for transmitting infection.

In the Uganda trial there were no lepromatous cases in the placebo group. The lepromatous rate at the beginning of the Burma trial was predicted to be 23 per cent³ but the actual rate in the placebo group was only 3 per cent. The main reason for this very low rate was

the rapid decline in case detection because of chemotherapy with dapsone, with a disproportionate fall in the lepromatous rate⁴. In contrast, in the BCG trial in Karamui, New Guinea, in a population of 5,000, chemotherapy was not available until 1967 and the incidence rate stayed constant at 5 per 1,000 per year from 1964 to 1971. There were 19 multibacillary cases during this period⁵ but only 4 from 1972 to 1979, this decline being due to the effect of chemotherapy which also lowered the incidence in the placebo group⁶.

C. L. Crawford

Department of Anatomy,
Charing Cross and Westminster
Medical School,
Fulham Palace Road,
London W6 8RF, UK

1. Jayaraman, K. S. *Nature* **356**, 373 (1992).
2. Convit, J. et al. *Lancet* **339**, 446–450 (1992).
3. Bechelli, L. M. et al. *Bull. WHO* **42**, 235–281 (1970).
4. Kyaw Lwin & Zuiderhoek, B. *Int. J. Lepr.* **43**, 125–128 (1975).
5. Scott, G. C., Russell, D. A., Boughton, C. R. & Vincin, D. R. *Int. J. Lepr.* **44**, 110–122 (1976).
6. Bagshawe, A. et al. *Bull. WHO* **67**, 389–399 (1989).

Natural gas

SIR — Tacitus, the Roman historian, describes in his *Annals* (Bk. XIII.57) a spectacular natural occurrence that took place in the neighbourhood of Colonia Agrippina in the year 58 AD. It is given here in Michael Grant's translation:

But a friendly tribe also, the Ubii, were overwhelmed by a sudden disaster. Flames bursting out of the ground devoured farmhouses, crops, and villages far and wide, right up to the walls of the recently founded settlement named after Agrippina. Neither rain nor river nor any other water could quench the fire. Finally, a few desperate and distraught peasants hurled rocks into the flames, and, as they subsided, advanced and fought them with clubs and other implements, as one would fight wild animals. Finally, they tore off their clothes and heaped them on. The oldest and dirtiest garments were most effective as extinguishers.

The phenomenon has all the characteristics of a leakage of natural gas from underground deposits, which may occur as a result of seismic activity, or natural erosion. The curious detail about the soiled garments is consistent with the way in which flaming natural gas leakages are choked off even today — by depriving the fire of oxygen. Tacitus describes the flames as covering a wide swathe of land around Colonia Agrippina (Cologne), on the western bank of the Rhine. Exploratory drilling may reveal the presence of natural gas deposits in this area.

Jan Sammer

PO Box 219,
Petersburg, New York, 12138, USA

Time for tempting the Nobel fates

The Nobel season has come round again, but there is no closed season on speculation about this or next year's winners, irreverent though it may seem.

THIS is the season when it is reasonable to begin guessing at who will be in this year's list of Nobel prizes, due out less than a month from now. Stockholm committee-members, delphic as ever, probably have a good idea already. Others should be aware of the risks. One is that of losing money by injudicious bets, even at long odds. Another is that even a good guess may seem to delay a merited award by several years, as if the committees were abreactive to suggestions; for three years after the mention of the name in this connection, this journal felt apologetic to César Milstein, fearing that it had spoiled his chances. Those mentioned in what follows, even indirectly, are asked to be forbearing.

The ground rules are important. Although Alfred Nobel's will specifies that the three science prizes (physics, chemistry and physiology and medicine) will be awarded for discoveries in the previous year, the committees are evidently not strictly bound by that consideration. And although the three named prizes do not formally include astronomy and the Earth sciences, the rubric is sufficiently elastic to have done some justice in these fields (but not, alas, to Hoyle). In passing, it is probable that Darwin, even had he been active later, would almost certainly not have qualified even under the most flexible interpretation; natural selection is certainly not medicine and is only at its most marginal physiology.

Techniques (as the example of chromatography shows) are not ruled out; given the present importance of the polymerase chain reaction (PCR) in amplifying small quantities (even single molecules) of nucleic acid polymers, that seems an obvious candidate. But Stockholm seems to prefer that the recipients of a single prize (the maximum is three) should be diverse in character; PCR will become even more awardable when an even more sensational use has been made of it than has yet been published.

Stockholm, like most journals, also seems to place particular value on a neat and well-rounded story. The now-successful hunt for the major cystic fibrosis gene has been exciting in itself, but will be irresistible in Stockholm when some physician has used what has been learned to cure patients. That prize cannot be long delayed. But there will be no prizes for, say, the nucleotide se-

quence of chromosome III of yeast; there are simply too many fingers in the pie, which is in any case a landmark rather than a discovery.

Much the same goes for the mechanism of gene control. Who could pick out three or fewer prizewinners from the army of people now working in the field? The best hope is that there will emerge some discovery about the mechanism of eukaryotic gene control suggesting unexpected continuity with the well-documented regulation of genes in bacteria. Much of the natural world is, in other words, unawardably complicated, although the regulation of the cell cycle may yet prove an exception. And the mechanism of the nervous system, unrecognized since the award to Hubel and Wiesel in 1981, is bound to come into its own again.

In the physical sciences, and chemistry in particular, there must be some place at Stockholm for C_{60} , now disarmingly called a novel allotrope of carbon. Its discovery has the classical virtues of seeming an historical necessity after the event. But it is probably now too soon to be celebrating. Stockholm may well prefer to wait until somebody has made an electronic device of it.

In that case, there is a case for doing something to applaud the importance of numerical techniques in understanding the ways in which molecules behave. Stockholm's reluctance to include machines among the recipients of its awards is understandable, but there is more than mere computer programming to the exploitation of, say, molecular dynamics in the understanding of complicated molecules. The difficulty would be in finding three or fewer people.

Straightforward physics is both more complicated and simpler. In one sense, there have been no sensations lately; the intermediate heavy bosons appeared on Stockholm's lists almost immediately after they were found, since when there has been endless argument about the nature of dark matter — and no top quark or Higgs boson. Anions are intriguing, but may not yet be prizeworthy.

A much better bet, perhaps the best this year, is linked with anions, whose statistical characteristics are usually defined by phase factors attached to their wave functions. It is now more than 15 years since Aharonov and Bohm pointed

out the consequences of the equations of quantum motion of an electron in an electromagnetic potential.

Wave functions that solve Schrödinger's equation are complex quantities, with real and imaginary parts, but because their physical significance (the likelihood that a particle will be at some point in space) rests exclusively on their product with their own complex conjugate, they can be multiplied by an arbitrary number of unit modulus, most generally the quantity $e^{i\theta}$, where i is the square root of -1 and θ is any angle. But can this phase have no physical significance?

Bohm's interest in this question seems to have been clearly motivated by his long spell as one of Einstein's close colleagues when Einstein had set out to show that Bohr's willingness to embrace indeterminacy as the cornerstone of quantum mechanics must be mistaken. So might not the apparently irrelevant phase of the wave function embody the hidden variables' Einstein sought? That is how it must have seemed.

Most dramatically, what is called the Aharonov-Bohm effect is best illustrated by the thought-experiment (since realized) with electrons fired from a source at a screen carrying two parallel slits — the Young's slit experiment in quantum mechanics — in which interference patterns of electron waves are normally produced. The trick is to put a narrow conducting solenoid behind the screen but parallel with the slits. Although the field may be altogether too small to influence the electrons directly, their phase is affected by the direction in which they move relative to the magnetic field — and the interference pattern is shifted accordingly.

Stockholm need not be impressed, of course. Useful and interesting though the idea has proved to be, it might be thought a predictable effect, even a curiosity. But suppose it is put together with M. V. Berry's generalization of the argument that phase factors crop up repeatedly when the energy of a system is a function of some independent variable (as is the strength of the magnetic field in the electron experiment)? Berry's argument, first published in 1984, cannot be mistaken for fun and games. There is a growing band of people wondering what it means for the foundations of quantum mechanics.

John Maddox

What goes up could come down

Jonathan Fink

WHEN Mount Pinatubo erupted last year, the huge amount of ash and gas that it pumped into the upper atmosphere blocked sunlight, lowering temperatures worldwide. At the same time, hot debris from the eruption cloud raced across the Philippine countryside as superheated pyroclastic flows that destroyed villages, crops and rainforests. Figuring out why explosive eruption columns sometimes rise and sometimes fall is the goal of a newly published laboratory study by A. W. Woods and C. P. Caulfield (*J. geophys. Res.* **97**, 6699–6712; 1992). By injecting buoyant mixtures of various organic solvents into tanks of water, they replicated the evolution of eruption plumes, defining the conditions associated with their growth and collapse. While confirming earlier theoretical models, their elegant experiments can also help us predict the dangers accompanying the most catastrophic of volcanic eruptions.

As the deaths of several volcanologists have shown, gaining a close-up view of a convulsing volcano rarely provides data worth the risk. Recent advances in remote sensing allow weather satellites, airborne radar, infrared spectrometers and gas 'sniffers' to obtain information safely from otherwise inaccessible eruption clouds. Numerical methods that use

supercomputers to model volcanic explosions are also increasingly popular although, with a few notable exceptions (G. A. Valentine, K. H. Wohletz & S. W. Kieffer *Bull. geol. Soc. Am.* **104**, 154–164; 1992), their value has come more from quantifying suspected processes than from identifying new ones.

The most common way to study eruptions is to map and interpret what they leave behind. For instance, sediments that rain down from high ash clouds form regular layers that drape evenly over hills and valleys, whereas deposits left by pyroclastic flows are more chaotic and tend preferentially to fill depressions. By comparing such sequences from different volcanoes, geologists have been able to reconstruct a whole catalogue of typical eruptive histories.

More and more volcanologists are supplementing these methods with laboratory simulations so they can systematically study otherwise unobservable phenomena. Recent experiments have replicated mantle hot spots, magma chambers, igneous intrusions, lava domes and pyroclastic flows using such mundane analogues as wax, clay, corn syrup and gelatin. Continuing this trend, Woods and Caulfield created a set-up that let them safely turn miniature explosive 'eruptions' on and off. They simulated

magma with buoyant combinations of methanol and ethylene glycol injected downwards into a tank of fresh water.

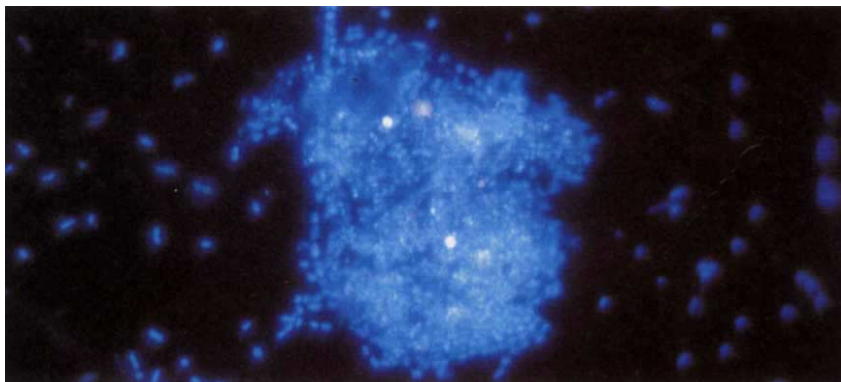
When mixed with enough water, these solutions had the unusual property of becoming more dense than either their organic constituents or the water itself. Normally the fluid jets would rise back up and spread across the top of the reservoir as soon as their momentum was dissipated. But injected forcefully enough to take in sufficient fresh water, they could become heavy and sink. The authors could precisely control whether the jets went down or up by varying their initial density and momentum, mimicking (in an inverted way) the transition from sustained eruption columns to collapsing 'fountains' that shed pyroclastic flows. Continued growth of the experimental plumes required only that they add enough water to remain denser than their surroundings.

Thermal effects complicate the analogous natural situation. Erupted ash is more dense than air, so its presence tends to promote column collapse. But these suspended fragments of magma also heat and expand the entrained air, giving it buoyancy. As a result, fallout of ash generally increases the bulk density and weakens the lift of an eruption cloud. Woods and Caulfield develop an analytical model that defines the critical eruption rate necessary to sustain columns in nature as well as in the laboratory. As was shown in calculations more than 10 years ago (L. Wilson, R. S. J. Sparks & G. P. L. Walker *Geophys. J. R. astr. Soc.* **63**, 117–148; 1980) tall eruption clouds are sustained by high ejection velocities and temperatures, low concentrations of ash, low particle densities and small-diameter vents. Subtle changes in any of these conditions can lead to column collapse and the formation of pyroclastic flows.

Although such experiments elegantly reproduce details of explosive volcanism, how reasonable is it to simulate giant eruption clouds in a glorified fish tank? As engineers have known for decades, properly constrained scale models are an inexpensive way to gain powerful insights into complex natural systems. They are best suited for identifying behavioural transformations like those between rising and collapsing eruption plumes. Also, many processes first recognized in the laboratory have later been confirmed by observations in nature.

For instance, Woods and Caulfield found that in some transitional experiments the sinking columns first stagnated and then gave off sporadic pulses of heavy fluid, like upside-down smoke signals. Although corresponding periodic clouds have not been described by volcanologists, their existence could explain why the deposits from some of the most

Decline and fall of marine snow



THE thriving colonies of bacteria (stained here with fluorescent dye) that inhabit 'marine snow' do not look after their hosts too well, it seems from the report of D. C. Smith *et al.* on page 139 of this issue. Indeed, much of the snow is rendered soluble by the enzymes released by the bacteria. This solves a puzzle of some years' standing. The downward flux of marine snow (aggregated particulate organic matter) rapidly shrinks with increasing depth, yet the microbial consumption of the snow is too small to account for its disappearance. But with solubilization rates 1–2 orders of magnitude greater than the rate of consumption it becomes clear why the aggregates last days not months. Keeping track of the various dissolved and particulate fluxes of organic and inorganic carbon in the oceans is one of the main tasks in unravelling the global carbon cycle, itself central to the question of atmospheric CO₂, so many will be relieved to see this particular puzzle cleared up.

R.P.

explosive eruptions in the geological record contain intimately mingled air-fall and pyroclastic-flow features. These pulsations might also be able to transport ash to the upper atmosphere more effectively than the central columns from the most violent eruptions because the latter tend to collapse before they can carry a great deal of material aloft.

Not only do laboratory models tell us about processes that might be hidden within ash clouds, they can also help us predict the consequences of eruptions more dangerous than any which scientists have yet witnessed. Pinatubo was estimated to produce the most energetic volcanic event of the twentieth century, but its size is dwarfed by prehistoric eruptions from such places as Crater Lake (United States), Taupo (New Zealand) and Cerro Galan (Argentina). Nearly all that we know about these

larger cataclysms comes from analyses of their sediments; we have little idea of the details of their evolution.

Over the past decade, measurements of volcano deformation have raised fears that comparable catastrophes might be imminent in several populated areas, including the Phlegrean Fields near Naples, the Mammoth Lakes resort area in California and Rabaul, Indonesia. Were an eruption of this scale to occur tomorrow, our ability to predict its effects on global climate and on the surrounding manmade and natural environment would be greatly improved by extrapolations from experiments like those of Woods and Caulfield. □

Jonathan Fink is currently with the Petrology and Geochemistry Program, National Science Foundation, 1800 G Street NW, Washington DC 20550, USA.

NEUROTRANSMITTERS

ATP joins the fast lane

C. D. Benham

SIGNALLING between neurons in the central nervous system can broadly be divided into slow and fast transmission. A bewildering array of putative transmitters exert their slow modulatory effects over a period of seconds, but the group of fast transmitters acting on a millisecond timescale forms a much more select club. So it is something of a special event when a new member is announced. A recent letter to *Nature*¹ and another from Edwards, Gibb and Colquhoun on page 144 of this issue² report the first convincing evidence that the ATP molecule can act as a fast excitatory transmitter at synapses between neurons, thereby joining the exclusive ranks of glutamate, acetylcholine and serotonin.

The idea that ATP can function as a neurotransmitter is not new. A weighty body of work has convinced all but the most intransigent sceptics that ATP is a fast cotransmitter at the sympathetic neuroeffector junctions of many vascular and visceral smooth muscles³. Within the nervous system itself, however, evidence for a transmitter role for ATP has been much more circumstantial. A variety of central and peripheral neurons are excited by ATP and the stimulus-evoked release of ATP has been demonstrated, but until now there has been no clear evidence that synaptic currents are mediated by ATP.

One hold-up to progress was the lack of a good antagonist to define the action of the endogenous transmitter. Also, there was a need to improve the resolution of methods for recording synaptic

currents in the central nervous system so that quantitative experiments could be done. The work of Evans *et al.*¹ has shown that suramin, a recognized ATP-receptor antagonist that lacks selectivity for P₂ purinoreceptor subtypes, is nevertheless an effective tool for discriminating between fast P_{2X} responses and those of the other fast transmitter candidates. In the coeliac ganglion, a waystation in the visceral nervous system, excitatory junction currents can be mimicked by application of exogenous ATP. The time course, reversal potential and current-voltage relationship of the synaptic current is reproduced by ATP. Most importantly, both endogenous transmitter and ATP-evoked currents are blocked competitively by suramin to the same extent, whereas currents evoked by the known fast transmitters were unaffected. The authors exploited the relative simplicity of this model synapse to complete an elegant pharmacological study. It is clear that ATP or a very similar molecule must be the transmitter at these neuro-neuronal synapses in culture.

Armed with this validated antagonist, Edwards *et al.*² have now identified ATP-mediated synaptic currents in a slice preparation from the central nervous system. Synapses using ATP are thus present in the brain and, importantly, in a slice preparation in which the potential for artefacts introduced by culturing cells is not a concern. They used the technique of whole-cell recording from brain slices with patch-clamp electrodes⁴, which enables synaptic currents to be recorded with higher resolu-

tion than with conventional intracellular electrodes. Fast synaptic currents were evoked that were not blocked by the expected transmitter antagonists, but were blocked by application of suramin and the desensitizing purinergic agonist α,β -methylene ATP, as were the miniatures seen when the transmitter was spontaneously released. Although the pharmacological characterization is less extensive, taken together with the coeliac ganglion work, the identification is convincing.

How widespread is central purinergic innervation? Excitatory actions of ATP in the central nervous system are well-documented. Whole-cell recording in slices using suramin as an antagonist should clarify how extensive the use of this transmitter system is. This search would be helped considerably by the cloning of the putative receptor-channel complex. In spite of their close functional similarities, probes derived from other ligand-gated cation channels that hybridize under conditions of low stringency have so far failed to pull out the ATP channel, by design or accident. What we do know is that the distribution of ATP release sites in the central nervous system most closely follows that of acetylcholine, suggesting that it may be a cotransmitter at cholinergic synapses, as in the periphery. Note that the medial habenula studied by Edwards *et al.*² is part of a recognized cholinergic pathway. Further comparison with peripheral systems indicates that it might be profitable to look at noradrenergic pathways.

There are several similarities between transduction systems coupled to purinergic and glutamatergic receptors. These agonists gate both ion channels and G-protein-coupled, inositol trisphosphate-mediated release of calcium ions from stores through different receptors. Both substances are almost ubiquitous inside cells, having important functions in cell metabolism. This has hindered acceptance of both molecules for the specialized role of an extracellular transmitter. It is likely, drawing on information from peripheral systems, that the ATP-gated currents are mediated by P_{2X}-like purinoreceptors directly

1. Evans, R. J., Derkach, V. & Surprenant, A. *Nature* **357**, 503–505 (1992).
2. Edwards, F. A., Gibb, A. J. & Colquhoun, D. *Nature*, **359**, 144–147 (1992).
3. Burnstock, G. & Kennedy, C. *Circ. Res.* **58**, 319–330 (1986).
4. Edwards, F. A., Konnerth, A., Sakmann, B. & Takahashi, T. *Pflügers Arch.* **414**, 600–612 (1989).
5. Bean, B. P. *Trends pharmacol. Sci.* **13**, 87–90 (1992).
6. Benham, C. D. *J. Physiol., Lond.* **419**, 689–701 (1989).
7. Mayer, M. L. & Miller, R. J. *Trends pharmacol. Sci.* **11**, 254–260 (1990).
8. Iino, M., Ozawa, S. & Tsuzuki, K. *J. Physiol., Lond.* **424**, 151–165 (1990).
9. Mülle, C., Choquet, D., Korn, H. & Changeux, J.-P. *Neuron* **8**, 135–143 (1992).
10. Krishtal, O. A., Marchenko, S. M. & Obukhov, A. G. *Neuroscience* **27**, 995–1000 (1988).

coupled to an ion channel⁵ permeable to cations and with a significant Ca^{2+} permeability⁶. As with the glutamate receptor family, there are functionally distinct subtypes of purinergic channel⁵. The unique properties of the NMDA(*N*-methyl-D-aspartate)-gated channel have led to considerable interest in the functional role of Ca^{2+} influx through these channels⁷, but much less is known about the role of the Ca^{2+} influx that occurs through the AMPA ((2-aminomethyl) phenylacetic acid)⁸ and nicotinic⁹ cation channels that these purinergic channels more closely resemble. The ATP-gated channels show inward rectification whereas NMDA-gated channels are blocked by Mg^{2+} , limiting current flow at negative potentials. Consequently,

the ATP-gated channels will allow greatest Ca^{2+} influx at hyperpolarized potentials without the postsynaptic neuron firing action potentials. Stronger purinergic receptor excitation will lead to post-synaptic action potential firing with less shunting of the membrane conductance during the action potential¹⁰.

The significance of such subtle differences in the channel properties remains to be determined, but it is now clear that these purinergic channels have a role in fast signalling that is essential to all the rapid sensory, motor and cognitive functioning of the central nervous system. □

C. D. Benham is at SmithKline Beecham Pharmaceuticals, The Pinnacles, Harlow, Essex CM19 5AD, UK.

ACTIVE GALACTIC NUCLEI

Brilliance without black holes

Andy Robinson

A CONTROVERSIAL case is put in new work¹ that active galactic nuclei, the violent and luminous cores found in 10 per cent of galaxies, are not the product of activity near a black hole, as is commonly believed, but rather are the result of exceptionally vigorous but otherwise conventional star formation. The authors, Roberto Terlevich and colleagues, base their case on the ease with which many of the spectral characteristics of active galactic nuclei can be modelled with what is already known about stellar processes, without recourse to more exotic astrophysical effects.

The light radiated by an active galactic nucleus (AGN) often exceeds that of the entire stellar population of the host galaxy, although the nucleus occupies only a tiny fraction of the volume. Most of the luminosity is radiated in a continuum which spans the electromagnetic spectrum, with about equal power per decade of frequency, from the far infrared to γ -ray energies. The optical-ultraviolet spectrum is also characterized by prominent emission lines of H, He and various ions of the more abundant heavier elements, notably C, N, O and Fe. In the more luminous AGNs (including Seyfert 1 galaxies and quasi-stellar objects or QSOs), spectral lines arising from permitted and semi-forbidden transitions often have enormously broad wings, indicating velocities in the emitting gas of up to 10^4 km s^{-1} .

Conventional wisdom has it that the enormous power of AGNs cannot come from the processes associated with normal stellar evolution, and instead a standard model has evolved in which the ultimate source of energy is a massive black hole at the Galactic Centre which

accretes gas and dust from its surroundings. According to this model, a fraction (perhaps 10 per cent) of the rest mass of the accreted material is converted into radiant energy and relativistic particles. The observed continuum radiation is produced partly by synchrotron emission and other non-thermal processes in relativistic plasma surrounding the black hole and partly by thermal emission from the hot inner regions of an accretion disk. The broad emission lines are emitted by dense gas clouds, moving in the gravitational field of the black hole, which are photoionized by radiation from the continuum source.

Earlier attempts to explain AGNs as dense star clusters and multiple supernovae were abandoned in the face of evidence which appears to favour the massive-black-hole hypothesis. However, much of the strongest evidence, including relativistic bulk motions and strong, highly variable, non-thermal continuum emission, applies only to the minority (about 10 per cent) of AGNs which are powerful radio sources, while the subclass (known as 'blazars') which show these properties most clearly are even more rare.

The evidence favouring black holes is less compelling for the 'radio-quiet' majority of AGNs. Over the past few years, Terlevich and collaborators have developed an alternative to the massive-black-hole hypothesis in which these AGNs can be understood as violent bursts of star formation (starbursts) occurring in the relatively dense, chemically enriched gas which, they argue, will be present in the core of the host galaxy²⁻⁴. In their latest paper¹, they show that many of the observed spectral

features can be explained as a direct consequence of the evolution of the resulting star cluster which, after passing through several phases of activity⁴, culminates with massive stars exploding as supernovae. It is this phase which is identified with the activity characteristic of Seyfert 1 galaxies and QSOs.

The optical, ultraviolet and X-ray continua are produced by the supernova explosions and the interaction of the resulting debris (the supernova remnant) with a dense circumstellar gas shell previously formed by the stellar wind of the progenitor. Only a few supernovae a year are required to produce the observed luminosity of Seyfert 1 galaxies, but considerably higher rates, from tens to a few hundreds a year, must occur in the brighter QSOs. The broad emission lines are mainly produced in dense, fast-moving shells of cool material which form behind the shocks set up by the supernova blast wave and are then photoionized by thermal radiation emitted by the shocks. The emission-line intensities predicted by this model are in remarkably good agreement with those observed in AGNs.

Furthermore, a particularly attractive feature of this picture is that both the configuration of the line-emitting gas and the ionizing radiation field are determined entirely by the interaction of the supernova ejecta with the circumstellar gas. Consequently, the model employs far fewer free parameters (basically only one) than more conventional models for the broad-emission-line region, in which the ionizing continuum and many of the properties of the emitting gas are rather poorly constrained.

Further work by Terlevich *et al.*, reported at a recent conference, shows that correlated variability of the optical-ultraviolet continuum and the broad emission line intensities, similar to that often observed⁵ in Seyfert 1 galaxies, arises as a consequence of the dynamical evolution of the supernova remnant. Similarly, the optical continuum variability of QSOs can be reproduced by a random series of superimposed supernova events⁶. However, it is not yet clear if the evolving supernova remnant can produce realistic emission-line profiles; predictions for the profile shapes and their temporal evolution should prove an important means of testing the model.

Although Terlevich's group can claim a striking success in accounting for the optical-ultraviolet spectra of AGNs, the starburst model encounters difficulties in other wavebands. A particular problem is the very rapid X-ray variability (on timescales of 1,000 seconds or less) which is now known to occur in several objects. Although the mechanism responsible for the variability is not clearly

understood even in the black-hole model, the observed amplitudes and time-scales seem to require the conversion of rest-mass to radiation with an efficiency which is unlikely to be achieved in supernova explosions⁷.

Other problems are raised by recent reports of long-term periodicities in the X-ray variability⁸ and the discovery that, in the few objects which have been detected in γ -rays, a significant fraction of the total luminosity emerges at these energies⁹. It will be difficult to accommodate these results within the starburst model if further observations reveal them to be common properties among the AGN population.

Another limitation of the starburst model in its current form is that it provides no explanation for the powerful radio sources, including the vastly extended double sources which are among the most spectacular manifestations of nuclear activity. These features imply that the AGN has maintained a preferred axis over several million years, favouring a central engine consisting of a single, coherent object, rather than a random series of supernova explosions. Even though the optical-ultraviolet spectra of radio-loud AGN are often similar to those of their radio-quiet counterparts, and can therefore be explained by the starburst model, a massive black hole still seems to be a necessary component of these objects.

Whether they turn out to be right or not, Terlevich and colleagues, by showing that many important characteristics of AGNs can be explained by a starburst, have made an important contribution to the debate on these objects. Their work calls attention to the stellar component of AGNs and suggests that it may play a more important role than hitherto thought. It may not, after all, be such a straightforward matter to distinguish between activity due to starbursts and more exotic processes. □

Andy Robinson is in the Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK.

1. Terlevich, R., Tenorio-Tagle, G., Franco, J. & Melnick, J. *Mon. Not. R. astr. Soc.* **255**, 713–728 (1992).
2. Terlevich, R. & Melnick, J. *Mon. Not. R. astr. Soc.* **213**, 841–856 (1985).
3. Terlevich, R. & Melnick, J. *Nature* **333**, 239–240 (1988).
4. Terlevich, R. in *Evolutionary Phenomena in Galaxies* (eds Beckman, J. E. & Pagel, B. E. J.) 149–158 (Cambridge University Press, 1989).
5. Clavel, J. *et al. Astrophys. J.* **366**, 64–81 (1991).
6. Terlevich, R. in *Structure and Dynamics of the Interstellar Medium* (eds Tenorio-Tagle, G., Moles, M. & Melnick, J.) 343–350 (Springer, Berlin, 1989).
7. Fabian, A. C. in *Testing the AGN Paradigm* (eds Holt, S. S., Neff, S. G. & Urry, C. M.) 657–667 (American Institute of Physics, New York, 1992).
8. Done, C. *et al.* in *Testing the AGN Paradigm* (eds Holt, S. S., Neff, S. G. & Urry, C. M.) 83–92 (American Institute of Physics, New York, 1992).
9. Gehrels, N. & Cheung, C. in *Testing the AGN Paradigm* (eds Holt, S. S., Neff, S. G. & Urry, C. M.) 348–355 (American Institute of Physics, New York, 1992).

Hunt for the magnetoreceptor

Rüdiger Wehner

ALTHOUGH the Chinese used magnetic compasses in the second century¹, it was not until the nineteenth that the Earth came to be regarded as a huge magnetic dipole² which created a three-dimensional geomagnetic field that might provide directional information to animal navigators³ (see figure). For example, when the entomologist J. H. Fabre was startled by the amazing homing abilities of bees and wasps that he had displaced over several miles, Darwin persuaded him to attach small magnets to his bees to try to manipulate their behaviour⁴; the bees, however, responded to this indignity with erratic flight manoeuvres and crash landings. Even now, the sensory mechanism by which an animal acquires its own magnetic compass remains elusive. But on page 142 of this issue⁵, Phillips and Borland describe experiments with newts which indicate that this magnetic sense could be dependent on light, and they suggest that it is the animals' photoreceptors that are responsible for picking up geomagnetic information.

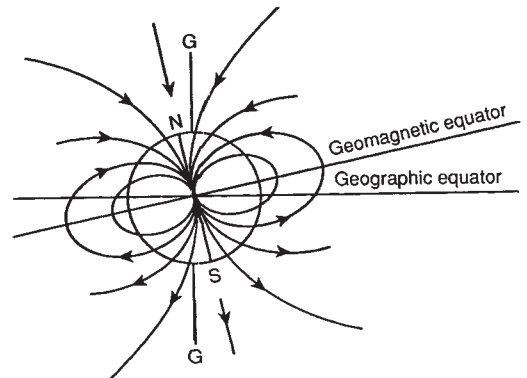
The most intensive research into the magnetic cues that might influence animal orientation has focused on birds, although it has been claimed that bees and termites also respond to the Earth's magnetic field⁶. Experiments in which nocturnal migrants like robins and warblers were exposed to artificial magnetic fields, or in which pigeons wearing bar magnets or small Helmholtz coils were released in 'unknown' territory, led to two conclusions: first, birds are able to use the geomagnetic field as a compass to find and maintain direction^{7,8}; and second, the avian compass is an inclina-

tion compass, referring to the dip angle rather than the polarity of the magnetic field, and so can distinguish between poleward and equatorward rather than between north and south⁹. Furthermore, it has been argued^{10,11} that besides exploiting the Earth's magnetic field as a compass, the homing pigeon can determine its geographical position by using a 'magnetic map' based on a bicoordinate grid in which at least one of the two variables is of a magnetic kind. Nevertheless, in spite of the plethora of data now available, the upshot of all this experimental inference is uncertainty. The data are extremely noisy; and the different types of experiments have rarely been repeated independently (and if they were, they more than once failed to confirm the previous results¹²).

More discouraging, however, is the fact that the search for the animal's magnetoreceptor has gone completely unrewarded. As the motion of any electric charge, and hence of biological tissue as well, is influenced by magnetic fields, magnetoreception could be based on intraorganismic ionic currents. Sharks and rays take advantage of such induced electromagnetic forces by using their electroreceptors for detecting magnetic fields¹³ but, notwithstanding some theoretical considerations¹⁴, in all other organisms the search for sensory mechanisms of magnetoreception has focused on direct (magnetic) rather than indirect (electric) responses to magnetic stimuli.

Here, the richest vein of discovery was neither in insects nor in birds, but in prokaryotes¹⁵. Certain aquatic bacteria such as *Aquaspirillum* exhibit magneto-

Owing to dynamo processes in the rotating Earth's fluid iron core, our planet is endowed with a dipolar magnetic field, in which the dipole axis roughly coincides with the axis of rotation, i.e. the geographic axis (G–G) of the Earth. The field lines leave the surface of the Earth at a pole (S) close to the southern geographic pole, and re-enter it at a pole (N) near the northern geographic pole. Hence, the angle of inclination formed by the field lines with the horizontal varies from 0° (at the geomagnetic equator) to 90° (at the geomagnetic poles). The total intensity of the geomagnetic field measured in tesla, T ($\text{N A}^{-1} \text{m}^{-1}$), decreases from the poles, where it is about $6 \times 10^{-5} \text{ T}$, to half that value at the equator. Based on behavioural experiments it has been argued that migratory birds derive directional information from a magnetic inclination compass^{8,9} and that they might even be able to use gradients of either magnetic intensity or inclination for map-like navigation^{10,11}. The concept of the bird's magnetic map was developed in spite of the fact that the geomagnetic gradients of intensity and inclination are very shallow indeed (10^{-9} to 10^{-10} tesla per kilometre, and 10^{-3} to 10^{-2} degrees per kilometre, respectively).



Track event

ONE man's loss is another's gain, at least when the loss is 80,000 Korean sports shoes washed overboard a ship in heavy seas, and subsequently borne to the American coast by ocean currents. C. C. Ebbesmeyer and W. J. Ingraham Jr used the shoes' paths to test ideas about Pacific Ocean currents (*Eos* **73**, 361–365; 1992). On learning of the spill, from the *Hansa Carrier* south of the Aleutians in May 1991, the authors alerted beachcombers through personal contacts and local papers, who recovered over 1,300 shoes from as far afield as Queen Charlotte Island, Alaska, down to southern Oregon — and even Hawaii. The shoes appear to have been carried due east until approaching the North American coast, at which point they were swept north by the Alaska current or south by the California current. The authors hope that some may eventually turn up in Japan or Asia, and plead for colleagues to keep their eyes open.

Proton yield

How elastic is the proton? The question is important for those trying to work out the details of the forces between the three quarks that make up the nuclear particle. H. P. Morsch *et al.* have tackled it by bouncing α -particles off a liquid hydrogen target (*Phys. Rev. Lett.* **69**, 1336–1339; 1992). The α -particles are particularly effective at kicking the proton into an excited state (the 'Roper' resonance) that is generally accepted as involving radial vibrations. By careful analysis of the emerging α -particles' energy, the authors conclude that the compressibility, in the peculiar units of nuclear physics, is about 1.4 gigaelectronvolts — close to what would be hoped for, as the value should not be far different from the proton's mass, 1 GeV.

MHC at sea

In a provocative study in *Proceedings of the Royal Society* (**B249**, 163–171; 1992), R. W. Slade finds that some of the genes of the major histocompatibility complex (MHC) in the southern elephant seal show much less polymorphism than is the norm among terrestrial mammals. That has also been the result of earlier analyses of two species of cetacean. Slade argues that the data, preliminary as they are as yet, can be best explained by marine mammals being less subject than their terrestrial counterparts to pathogenic selection pressure on the MHC (molecules of which are central players in the immune response). Moreover, he points out that limited polymorphism in the MHC may be why marine mammals seem to be especially prone to devastating epidemics when a pathogen does attack a population.

tactic behaviour when they swim along the inclined lines of the Earth's magnetic field. They swim northwards in the Northern and southwards in the Southern Hemisphere, which means always downwards towards the mud. Each bacterium contains a chain of about twenty single-domain magnetite (Fe_3O_4) particles and so acts as a biological magnetic dipole, with either its north- or south-seeking pole directed forwards (in the Northern and Southern Hemisphere, respectively)¹⁶.

The discovery of the bacterial compass elicited a feverish search for biogenic magnetite in a variety of animal species. As more researchers hunted for concentrations of magnetite crystals in biological tissues, such crystals started turning up widely in animals ranging from bees to bats¹⁷. Many a scientist's hope was pinned on these discoveries, but until now there has not been even the slightest indication of a sensory mechanism by which this magnetic material could provide the animal with that elusive magnetic sense, be it compass or map.

Phillips and Borland⁵ attempt to take a fresh look at this mystery. After mechanoreceptors had been proposed to act as magnetoreceptors¹⁸, the authors now suggest that it could be the animal's photoreceptors that pick up geomagnetic information. This hypothesis was first proposed fifteen years ago by a theoretical physicist¹⁹. It implies that the light-evoked responses of specialized photoreceptors utilizing an electron spin resonance process are modulated by the Earth's magnetic field. The account of Phillips and Borland does not prove this hypothesis, but it might provide some indirect support for a light-dependent magnetic compass mechanism.

In this study the experimental animals are not birds, but newts, *Notophthalmus*, which were trained to orient in a particular direction with respect to the lines of force of the geomagnetic field. When tested under one of four artificial light alignments (magnetic north at geographic north, east, south or west), the newts kept their training directions con-

stant relative to the magnetic rather than to the geographic system of reference, but they selected different angles with respect to the magnetic field when they were illuminated by either short- (≤ 450 nm) or long-wavelength light (≥ 500 nm). When tested under 475-nm light, or in the dark, they were completely disorientated. The changes in magnetic compass orientation caused by changing the illumination in 25-nm steps from 450 to 475 to 500 nm are surprisingly clear and abrupt. In fact, they provide one of the most consistent responses to geomagnetic stimuli ever observed in any animal species.

Nevertheless, the route by which the authors attempt to infer the existence of a light-dependent magnetoreceptor is worryingly indirect. Photoreceptors sensitive to blue, green and yellow light ($\lambda_{\text{max}} = 430, 500$ and 580 nm, respectively) are found in amphibians²⁰, and physiological recording techniques are well established for the amphibian retina. Hence, the next logical — and feasible — step to take is to try to elicit modulations of photoreceptor responses by changes in magnetic fields within the intensity range of 10^{-6} to 10^{-5} tesla. A start in this direction may already be underway in birds, in which some single-unit recordings have recently revealed responses to magnetic stimuli in various parts of the visual system²¹.

The search for the magnetoreceptor has a long history of exciting preliminary results that have been difficult to replicate²². As the behavioural data collected by Phillips and Borland⁵ are very clear-cut indeed, there is hope that they are not just another such example and that they may eventually be underpinned by neurobiological investigation, which has been so successful in analysing the mechanisms behind animal orientation based on other unsuspected cues such as electric fields, ultrasound and polarized light²³. □

Rüdiger Wehner is in the Department of Zoology, University of Zürich, CH-8057 Zürich, Switzerland.

1. Needham, J. *Science and Civilization in China* Vol. 4 (Cambridge University Press, London, 1962).
2. Whittaker, E. A *History of the Theories of Aether and Electricity* (Nelson, London, 1951).
3. Viguier, C. *Rev. Phil. France Etrang.* **4**, 1–36 (1882).
4. Darwin, F. *The Life and Letters of Charles Darwin* Vol. 3 (Murray, London, 1888).
5. Phillips, J. B. & Borland, S. C. *Nature* **359**, 142–144 (1992).
6. Towne, W. F. & Gould, J. L. in *Magnetite Biomineralization and Magnetoreception in Organisms. A New Biomagnetism* (eds Kirschvink, J. L., Jones, D. S. & MacFadden, B. J.) 385–406 (Plenum, New York, 1985).
7. Keeton, W. T. *Proc. natn. Acad. Sci. U.S.A.* **68**, 102–106 (1971).
8. Wiltschko, W. & Wiltschko, R. *Curr. Ornithol.* **5**, 67–121 (1988).
9. Wiltschko, W. & Wiltschko, R. *Science* **176**, 62–64 (1972).
10. Gould, J. L. *Nature* **296**, 205–211 (1982).
11. Wallraff, H. G., Benvenuti, S., Papi, F. & Gould, J. L. *Nature* **300**, 293–294 (1982).
12. Moore, B. R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4907–4909 (1988).
13. Kalmijn, A. J. *Science* **218**, 916–918 (1982).
14. Yeagley, H. L. *J. appl. Phys.* **18**, 1035–1063 (1947).
15. Blakemore, R. P. *Science* **190**, 377–379 (1975).
16. Blakemore, R. P., Franke, R. B. & Kalmijn, A. J. *Nature* **286**, 384–385 (1981).
17. Kirschvink, J. L., Jones, D. S. & MacFadden, B. J. (eds) *Magnetite Biomineralization and Magnetoreception in Organisms. A New Biomagnetism* (Plenum, New York, 1985).
18. Lindauer, M. & Martin, H. in *Animal Orientation and Navigation* (eds Galler, S. R. *et al.*) 559–567 (US Gov. Print. Off., Washington DC, 1972).
19. Leask, M. J. M. *Nature* **267**, 144–145 (1977).
20. Roth, G. *Visual Behavior in Salamanders* (Springer, Berlin, 1987).
21. Beason, R. C. & Semm, P. in *Orientation in Birds* (ed. Berthold, P.) 106–127 (Birkhäuser, Basel, 1991).
22. Griffin, D. R. *Experientia* **46**, 335–336 (1990).
23. Laughlin, S. B. (ed.) *J. exp. Biol.* **146**, 1–334 (1989).

How does variation count?

Alice S. Huang and John M. Coffin

WHAT do virologists, evolutionary biologists, population geneticists, mathematical modellers, epidemiologists and physicians have in common? At a recent EMBO workshop* an area of overlap emerged that could grow into a powerful scientific interface. New approaches were reported for analysing viral disease processes and evolutionary relationships, in which computer simulations were combined with the wealth of molecular data being generated by the sequencing of viral genomes. Complex but manipulatable experimental viral systems were used to test theories of population dynamics. And, almost as an aside, it was firmly established that all RNA viruses, not only human immunodeficiency virus, mutate — but do not necessarily evolve — at a high rate, and that recombination is relatively common among RNA viruses.

Ever since von Magnus described changes in influenza virus populations almost 40 years ago, investigators have sought the cause of these changes. They have wondered whether such variations, so easily detected in cell culture systems,

could account for altered virulence in the infected host. Sparked by what seems to be a remarkable genetic plasticity among human immunodeficiency virus and the annual emergence of threatening new influenza virus strains, a closer inspection was made of what is currently known about mutation rates for viruses. Rates of mutation for RNA genomes are now estimated to be 1,000-fold higher than for DNA genomes (J. Drake, National Institute of Environmental Health Sciences). Furthermore, the rates are highly variable from site to site in the same genome (T. Kunkel, National Institute of Environmental Health Sciences; B. Preston, Rutgers; M. Ricchetti, Pasteur Institute). This micro-heterogeneity leads to a population of viral genomes known as quasispecies, which are closely related but different from one another. The proportions of these quasispecies change, depending on both intrinsic error rates in polymerases and selective forces that may vary according to the host environment. From the composition of different quasispecies, viral replication rates, fitness selection and mutation rates can all be estimated in that virus population

(M. Nowak, University of Oxford).

Using improved molecular technology, particularly the ready amplification of nucleic acid sequences by the polymerase chain reaction, mutational heterogeneity is easily detected among different specimens taken at the same time and place from different patients, and in sequential specimens taken from the same patient. Although genetic variation may itself be important for pathogenesis, the immune system provides an important selective force (R. Zinkernagel, University of Zurich; R. Phillips, John Radcliffe Hospital, Oxford; S. Carpenter, Iowa State University; A. Leigh-Brown, University of Edinburgh). Disappointingly, there were no clear-cut correlations between the degree of heterogeneity and the course of disease. It may well be that genetic change among viruses is intrinsic and the number of mutations simply related to the number of replication cycles (S. Wain-Hobson, Pasteur Institute), resulting in a genetically diverse population capable of responding to very small selective forces.

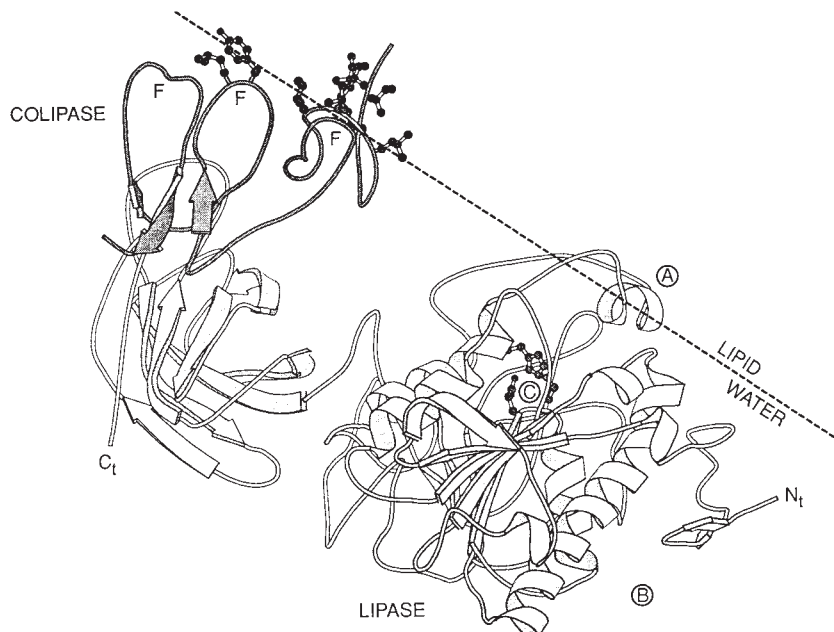
Besides the heterogeneity resulting from single base mutations, a range of strange and wonderful variations among viruses was presented. These included satellite viruses of plant and animal viruses, where there is no commonality of sequence between the parasite and its

*EMBO Workshop on Variation and Molecular Evolution of Viruses, Grignon, France, 7–11 July 1992.

Lifting the lid on enzymatic breakdown of lipids

LIPASES are enzymes that hydrolyse triacylglycerol lipids into free fatty acids and glycerol — a crucial step in breaking down ingested fats in the gut. The pancreatic lipase enzyme's activity is greatly increased at the lipid–water interface, when a short amphipathic helix, or lid (A), in the amino-terminal domain of the lipase (B), pops open and exposes the active site (C) of the enzyme to its lipid substrate (a phenomenon known as interfacial activation). The interaction of the lipase enzyme with the lipid interface is very sensitive to the chemical nature of the lipid surface; in the presence of amphiphiles such as bile salts and phospholipids, the binding of the lipase to the lipid surface requires a small protein cofactor, colipase. The complex of these two molecules reveals that colipase binds to the carboxy-terminal domain of the pancreatic lipase molecule (see the paper by H. van Tilbeurgh *et al.* on page 159 of this issue). The colipase's three hydrophobic fingers

(F), here ball-and-stick groups, are sample hydrophobic amino-acid side chains that point away from the lipase molecule but are on the same face of the complex as the lid that covers the active site of the lipase molecule. The binding of lipase, but not colipase, to the lipid–water interface is inhibited by amphiphiles (like bile salts). In the



presence of such inhibitors, then, the colipase cofactor may bring the larger lipase enzyme into close contact with the interface and allow opening of the lid to expose the active site, so facilitating the enzymatic breakdown of lipids. (Figure adapted from P. J. Kraulis *J. appl. Crystallogr.* 24, 946–950; 1991.)

G. R.

helper (G. Bruening, University of California at Davis; W. Carman, Institute of Virology, Glasgow); viroids, which are small RNA replicons, containing duplicated sequences (P. Keese, CSIRO, Canberra); illegitimate recombinants among viruses containing either RNA or DNA genomes (H. Temin, University of Wisconsin; M. Lai, University of Southern California; G. Smith, University of Oxford); hypermutation of G to A in retrovirus genomes (J.-P. Vartanian, Pasteur Institute); nonreplicative modification of nucleotides in RNA of paramyxoviruses (M. A. Billeter, University of Zurich); and the insertion of host sequences into the RNA of a flavivirus, bovine viral diarrhoea virus. The insertion introduces a change in a cleavage site, producing a new protein that changes the course of disease in persistently infected cows to death from mucosal peeling within a few hours (N. Tautz, Federal Research Centre for Virus Disease of Animals, Tübingen). Results like these promise that differences in disease potential may indeed be due to heterogeneity in viral populations in which chunks of genetic material are being inserted, deleted or otherwise altered. The challenge will be to dissect out population changes in hosts during the evolving pathogenic events and to relate attenuation or virulence to each of the particular variations found in these viruses.

Hundreds of sequences from viral isolates and more distantly related viruses are grist to the mill for an evolutionary biologist or mathematical modeller. Different algorithms for analysing sequence relationships among viruses yield evolutionary trees relating one isolate to another within families of viruses. These trees relate isolates from individuals, obtained during an epidemic or during pandemics evolved over a century, with isolates from related host species that span from thousands to millions of years (A. Leigh-Brown; W. M. Fitch, University of California at Irvine; M. Eigen, Karl-Friedrich-Bonhoeffer Institute). Gaining recognition is the application of pattern recognition to sequences in search of similar catalytic functions among diverse proteins (M. McClure, University of California at Irvine). The problem of applying a timescale to such trees is severe. Molecular clocks based on neutral theory are inappropriate for viral quasispecies: RNA viruses can be genetically very stable despite high mutation rates (E. Domingo, Universidad Autonoma de Madrid) and the rate of change of influenza virus is quite different in different hosts (W. M. Fitch; O. Gorman, St Jude's Hospital). Observations like these call into question recent estimates of the age of human immunodeficiency virus and its relatives.

One group is even attempting, so far without success, to approach this issue by analysis of monkey bones from Egyptian tombs for simian immunodeficiency virus-like sequences (A. van der Kuyl, University of Amsterdam).

To population theorists, quasispecies, recombinations and deletion mutants, coupled with the relatively short generation times of viruses, offer interesting systems to test hypotheses *vis à vis* real data. For instance, there is Muller's ratchet, a hypothesis that states that mutation will move a small population inexorably more distant from the most fit genotype; to ensure survival, such driven populations gain stability through increasing their population size or through recombination. Continuous passaging of an RNA phage under nonselective conditions and measuring mutants and recombinants led to complete support for this hypothesis (L. Chao, University of Maryland).

Another successful use of mathematical modelling was based on the predator-prey relationship described for a defective interfering particle and its helper standard virus. Rather than regular cycling between defective interfering particles and standard virus, there was chaotic unpredictability even in a deterministic model (T. B. L. Kirkwood, National Institute for Medical Research, Mill Hill). When mutations were added to the model, computer-simulated viral passages showed that new defective interfering particles periodically emerged and drove the level of replication-competent virus so low that there was continual pressure for viruses to evolve and escape interference.

Unlike most evolutionary studies, there is immediate and important practical value to understanding the evolution of viruses. Close examination of examples of recently emergent viruses implies that the key is altered conditions permitting efficient spread of virus in its new host rather than the initial infection events themselves (S. Morse, Rockefeller University). Armed with knowledge of evolutionary processes, perhaps we can head off the next pandemic. Mathematical modelling based firmly on experimental results has considerable potential for understanding viral diseases, and evolving populations and virus systems will provide tools for mathematical biologists to test their theories. Certainly more workshops at this interface and other multidisciplinary interfaces are needed. □

Alice S. Huang is at New York University, New York, New York 10003, USA. John M. Coffin is in the Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts 02111, USA.

DAEDALUS

Know thyself!

THE victorian art of phrenology divided the brain into compartments. One part dealt with hope, another with constructiveness, another with language and so on. The strength of each faculty could be judged by the enlargement of the skull at that point. Feeling their clients' skulls with sensitive fingers, popular phrenologists provided vocational guidance, character assessment and psychiatric diagnosis. They even tried to detect an overactive faculty from the warmth of the skull over the relevant region. Nowadays, these skills have practically died out.

Daedalus plans to revive them. He points out that modern techniques such as computed X-ray tomography give us direct access to the shape and size of the brain. Nuclear magnetic scanning, and especially positron-emission imaging, can even map its metabolic activity. So DREADCO technicians are now setting up a laboratory of Positron-Emission Phrenology to test the old ideas with modern rigour and explore new ones.

Volunteers for the new study will simply lie in the PEP machine and respond to requests to think about a particular topic or exercise some specific emotion. The site of sudden metabolic activity, or raised concentration of the implicated neurotransmitter, will then be displayed on the machine's brain map. For example, classical phrenology located the sexual instinct at the back of the head, rather below the level of the ears. Since (it is claimed) the average male thinks about sex every ten minutes, the metabolism of this region in males should show a ten-minute periodicity.

Even more revealing, volunteers will be allowed to watch their own brain display. Simply by letting his thoughts range, a subject will quickly discover what thoughts or mental efforts give rise to what activity where. By rapid biofeedback exploration, he will soon map out his own personal phrenology. He should then be able to learn how to stimulate particular regions of his brain, and to think himself into little-used areas. Those who normally rely on the left side of their brains will discover how to exercise their creative right side. Those out of touch with their feelings, or psychologically burdened with repressed memories, may be able to explore their brain map until they have tracked them down and inspected them. Those tortured by obsessive thoughts will find out at last where they lurk in the brain. They may then learn how to avoid stimulating that region, or how to swamp it with a mass of numbing platitudes. A few minutes on the machine could save you years on the couch.

David Jones

Self-cleaning of the Gulf

SIR — It has been calculated that the Arabian Gulf is charged yearly with 160,000 tons of oil arising from legal as well as illegal activities¹. Nevertheless, the deliberate release of oil by the Iraqi forces during the occupation of Kuwait (2 August 1990–26 February 1991) has left the Gulf environment confronted with the greatest single oil pollution event known. On 19 January 1990, the Iraqi forces released into the Gulf from Mina Al-Ahmadi oil terminal about 500,000 tons of crude oil. This oil has drifted from the Kuwaiti coasts, and has in part severely polluted about 770 km of the Saudi Arabian coasts. Thus, for example, the coast of Jubail, a Saudi research station about 200 km south of the Kuwaiti borders, has become covered with up to 20 kg m⁻² oil, and the free rock surfaces with up to 1.4 kg m⁻² (ref. 2). This research station falls within the coast-line sector reported to be most polluted with oil³.

Weekly inspection of the Jubail coast between May and December 1991 has revealed no substantial improvement except for one interesting phenomenon, the intensive appearance of blue-green microbial mats over oil layers. These mats now cover the regularly inundated lower part of the oiled intertidal zone. Strikingly, the microbial mats are intimately associated with oil, so that oil-free areas are also free of such mats. Subsequently, the mats with the strongly adhering oil layers are subjected to tearing and dissection into irregular scales of varying dimensions, about 10–30 cm in diameter. The whole area appears as if it has just been subjected to farming management (see figure).

All forms of higher life on the oiled beaches have been seriously affected by the oil pollution. We have observed that animals such as *Ocypode* and *Cleistostrima*, once common inhabitants, are absent or dead, and their underground channels in the coast are full of oil. It appears that the natural cyanobacterial grazers have been severely affected by the oil. Consequently, the cyanobacterial mats, which are not nutrient-limited, have had the chance to flourish. In this context, Readman *et al.*³ noted that subtidal sediments were poor in hydrocarbons, and the near-shore animals, for example, molluscs and fish, survived.

We believe that the intensive growth of blue-green microbial mats on the top of oil in Saudi coasts is the first sign

of self-cleaning activity in this polluted region. These microbial associations appear to be the only living things in this environment.

We have analysed the composition of the mats and find that they are associations of photosynthetic and nonphotosynthetic microorganisms. In most localities, one filamentous cyanobacterium, *Microcoleus* sp., is predominant, making up about 65% of the biomass volume. The mats also contain *Spirulina*, various diatoms and small numbers of many



Scaling of microbial mats on oily coastline

other unidentified green microorganisms. In a few regions, the predominant microorganism was another chromogenic (red), unicellular cyanobacterium which we have not yet identified.

One characteristic of these predominant cyanobacteria is that they produce large amounts of mucilage in which the cells and filaments are embedded. Whether these photosynthetic organisms can degrade oil is still under investigation. Only a little is known about the ability of photosynthetic microorganisms to biodegrade hydrocarbons⁴. However, we have found embedded in the mucilage, intimately associated with the cyanobacteria, up to one million cells per gram of fresh mat of organotrophic bacteria capable of utilizing crude oil and individual *n*-alkanes as sole sources of carbon and energy. There is no information in the literature on the occurrence of such mats before the oil catastrophe. However, cyanobacterial mats containing *Microcoleus* sp. as a minor constituent have already been recorded in the unpolluted coast of Abu Dhabi⁵, yet there is no mention of oil-degrading bacteria in such mats.

The presence of such large numbers of oil-degrading bacteria immobilized within the mats suggests that oil degradation is being stimulated by nitrogen leaching from the cyanobacteria. Immobilizing oil-degrading bacteria within the cyano-

bacterial mucilage protects them from being washed out to the open sea and, in addition, provides them with oxygen from the photosynthetic partner. Oxygen is essential for hydrocarbon biodegradation. Laboratory studies have shown that calcium alginate-immobilized microorganisms maintain their alkane-degrading activity^{6,7}, so it is possible that complex microbial associations can efficiently biodegrade the oil pollutants.

Finally, the mats we describe here have not been found to be associated with earlier oil spills elsewhere, and appear to be unique to the extreme environment of the Gulf.

Naser Sorkhoh

Redha Al-Hasan

Samir Radwan

Department of Botany and Microbiology,
University of Kuwait,
PO Box 5969,

Safat 13060, Kuwait

Thomas Höpner

Institute for Chemistry and Biology of the
Marine Environment (ICBM),
University of Oldenburg,
Oldenburg, Germany

1. Oostdam, B. L. *Mar. Pollut. Bull.* **11**, 138 (1980).

2. Höpner, Th., Berthe-Corti, L., Harder, H. & Michaelsen, M. *Erdöl Kohle* **44**, 197 (1991).

3. Readman, J. W. *et al. Nature* **358**, 662 (1992).

4. Cerniglia, C. E., Gibson, D. T. & van Baalen, C. J. *gen. Microbiol.* **116**, 485, 495 (1980).

5. Golubic, S. in *Environmental Evolution: Effects of the Origin and Evolution of Life on Planet Earth* (ed. Margulis, L.) 103–130 (MIT, Cambridge, Massachusetts, 1992).

6. El-Aassar, S. A., Omar, S. H. & Rehm, H.-J. *Appl. Microbiol. Biotech.* **29**, 442 (1988).

7. Heinrich, M. & Rehm, H. J. *Eur. J. appl. Microbiol. Biotech.* **11**, 139 (1981).

Patch production by cattle

SIR — In his News and Views about herbivore influences on grassland heterogeneity, Moore¹ implies a relative poverty in “the patchy environment created by the leaky cow”. However, he does not consider the East African grasslands, which provide one of the richest examples of spatiotemporal heterogeneity created by herbivores, and support in turn a unique density and diversity of grazing mammals. The 45,000-km² Serengeti Ecological Unit has 35 wild ungulate species, of which 15 coexist in superficially similar niches. In addition, pastoralists like the Maasai have grazed their herds here for more than 2,000 years. In 1960, the Maasai were excluded from the 15,000-km² Serengeti National Park, but they continue to use that part of the Serengeti Ecological Unit with the highest seasonal herbivore density and diversity, the 8,000-km² Ngorongoro Conservation Area. Grassland heterogeneity in this joint land use area still owes as much or more to the

patchy environment created by the leaky cow as it does to wild herbivores.

We have analysed² the contribution of the Maasai and their herds to ecosystem dynamics, including patch production, in relation to the part played by wild grazers. Wildebeest have come to dominate the herbivore community in the Serengeti Ecological Unit since the exclusion of the Maasai from the national park. The wildebeest seek out high nutrient concentration areas in which to feed but ruminate, defaecate and urinate in short-grass habitats different from the areas they graze³. However, the resulting processes of nutrient stripping and concentration are taken to far greater extremes by Maasai herds which graze in selected pastures, but ruminate, defaecate and urinate in night enclosures. The resulting high-nutrient patches revert to grazing land after a few weeks (for temporary wet season camps), years (in permanent arid zone settlements) or decades (in wetter highland zone permanent settlements)².

The overall structure and species diversity of these Serengeti mosaic grasslands are remarkably resilient while grazing is maintained. Belsky⁴ has shown the rapid recovery of both species diversity and spatial structure following a variety of natural and experimental artificial disturbances. In contrast to other conservation areas, the Serengeti grasslands resist invasion by introduced species². However, enclosure of grazers leads to loss of productivity and diversity.

Patch production by cattle can go too far. Tolsma *et al.*⁵ have shown that nutrient stripping by livestock generates phosphate and other deficiencies in intensively ranches Botswana rangelands, while concentration of nutrients around livestock watering holes leads to the growth of toxic plants. However, in Maasailand, 20 years of attempts to intensify livestock production have failed to change standard measures of the ecology of livestock production⁶. It is just as well for the conservation status of one of the world's richest and most diverse grassland ecosystems that the leaky Maasai cow retains its traditional role.

Katherine Homewood
Department of Anthropology,
University College,
Gower Street,
London WC1E 6BT, UK

"Surface tension" in the lungs

SIR — In a letter to *Nature* in 1955, Pattle¹ noted that the foam in cases of acute pulmonary oedema was peculiar in that the bubbles were extremely small and very stable. He used two phrases to explain the phenomenon that he had so accurately observed. First, "The surface tension of the lung bubbles is therefore zero"; and subsequently "It is thus evident that the alveoli are lined with an insoluble protein which can abolish the tension of the alveolar surface." Thus the word 'tension' carries two distinct meanings; the first (dynes per cm) as a manifestation of surface free energy (ergs per cm²) giving rise to the spontaneous contraction of liquids, a consequence of the unequal pull experienced by molecules in the interface between two phases having unequal concentrations; and the second, where a solid surface is too rigid to be distorted by surface free energies.

Recent claims to have achieved zero^{2,3} or near-zero⁴ surface tension at air/water interfaces are reminders that a body of lung physiologists and neonatologists, following up Pattle's observations, failed to distinguish between the two, quite different, implications of the term 'tension', with the result that a custom has prevailed in which all measurements from whatever device are presumed to identify with the 'tension' of the surface free energy kind.

It is remarkable that these claims, persistent over the past 30 years and pertaining to lung surfactant(s) placed on water in a Langmuir trough or within tethered bubbles or, by implication, within the alveolar spaces of a lung, had not been challenged until recently⁵. The claims are absurd as zero surface tension is tantamount to zero surface free energy, which in turn implies absence of interface. Yet no author has ever described the spontaneous demise of (standing) water into vapour following the application of lung surfactant. The surface tension of a water/vapour interface approaches zero only at the critical temperature (374.1 °C), at which point the concentration of liquid water is equal to the concentration of water vapour and a single phase is established.

Correct interpretation of the measurements is essential if one is to draw valid conclusions relating to lung mechanics, lung microanatomy and to the mechanism of action of lung surfactant. For example, the most common method of assaying lung surfactant is to measure the weight of the meniscus of water supported by a unit length of high-energy solid (for example, a heated

platinum plate). Any contamination of the plate by organic matter, whether phospholipid or protein, inevitably reduces the surface free energy of the plate to that below water and the weight of the meniscus falls. Such an event in no way reports upon the true surface tension of the sample solution and could reflect merely the affinity of a phospholipid or protein or both, whether as monomers or aggregated, to a clean metal surface².

Comprehension between physical and medical scientists would be resolved if the latter were to limit their reports to what they measure and not what they think they are measuring.

A. D. Bangham

17 High Green, Great Shelford,
Cambridge CB2 5EG, UK

1. Pattle, R. E. *Nature* **175**, 1125 (1955).
2. Cochran, C. G. & Revak, S. D. *Science* **254**, 566–568 (1991).
3. Schurch, S., Bachoven, H., Goerke, J. & Green, F. *Biochim. biophys. Acta* **1103**, 127–136 (1992).
4. Cockscht, A. M. & Possmeier, F. *Biochim. biophys. Acta* **1086**, 63–71 (1991).
5. Bangham, A. D. *Med. Sci. Res.* **19**, 795–799 (1991).

Fatal attraction to cyanobacteria?

SIR — Scavenging of shoreline deposits of neurotoxic benthic cyanobacteria by dogs has recently resulted in deaths at freshwater lochs (lakes) in Scotland. The first report in the scientific literature on animal deaths due to toxic cyanobacteria in lakes, by George Francis, appeared in *Nature* more than a century ago¹, and since then there have been many reports of cyanobacterial animal intoxications and the toxins responsible^{2,3}.

Until recently, the perceived circumstances of animal poisonings due to cyanobacterial toxins conformed to a general pattern. The buoyant growth habit of the widely encountered toxigenic, planktonic genera such as *Microcystis*, *Anabaena*, *Aphanizomenon*, *Nodularia* and some *Oscillatoria* species can result in scum formation in lakes and ponds during calm weather, so that an acutely toxic dose of cyanobacterial toxins can be presented to animals in a smaller volume than their daily water requirement. The animal deaths due to cyanobacterial toxins in Australia described by Francis¹ and elsewhere by many subsequent investigators have typically occurred after drinking from or near toxic scums^{2,3}.

Such poisonings have arisen under the following circumstances: if access to clear water is barred and animals are obliged to drink from regions containing toxic scum or blooms; or, as in the case of the deaths of 15 dogs and 20 sheep at Rutland Water, Leicestershire, in 1989

1. Moore, P. *Nature* **357**, 644 (1992).
2. Homewood, K. M. & Rodgers, W. A. *Maasailand Ecology* (Cambridge University Press, 1991).
3. McNaughton S. J. *Nature* **345**, 613–615 (1990).
4. Belsky A. J. *J. Ecol.* **74**, 1, 419–437, 11, 937–952 (1986).
5. Tolsma, D., Ernst, W. & Verwey, R. J. *appl. Ecol.* **24**, 991–1000 (1987).
6. Homewood, K. M. *Ecology of Food and Nutrition* (in the press).

(ref. 3), the entire lake margin is surrounded by toxic scum and it is not possible to wade out to open water without passing through the scum. Animal poisonings due to the ingestion of toxic planktonic cyanobacterial scum, or concentrated bloom, seem to occur as unavoidable consequences of contact with toxic planktonic scums and blooms. The habit that dogs have of licking adherent material from their coats has contributed to the ingestion of lethal doses of hepatotoxic *Nodularia spumigena* after swimming in water containing scum⁴. Whether the dogs or livestock were attracted to the planktonic cyanobacteria is unknown.

We have investigated dog poisonings that have occurred rapidly after ingestion of neurotoxic benthic cyanobacteria accumulated along the margins of Scottish lochs in 1990, 1991 (ref. 5) and in May and June this year. The most recent neurotoxicoses at lochs Awe and Avich (Argyll) have included two deaths and two cases in which the animals eventually recovered. As before, we have identified benthic *Oscillatoria* spp. and the cyanobacterial neurotoxin anatoxin-a in scums and dog stomach contents. When visiting the sites where the poisonings occurred, we found that the accumulations of toxic cyanobacteria were highly localized. For example, at Loch Awe the deposits extended along the shore for about 10 m, no more being encountered for more than 100 m. Why then did the poisonings occur when the animals had ready access to clear water for drinking and swimming?

The dogs at lochs Awe and Avich were accompanied by their owners, who saw the animals eat from the shoreline deposits of toxic cyanobacteria. These deposits partly occurred as slime and wet mats in the shallows and partly as dried crusts above the waterline, which had fallen during dry weather. Scavenging by dogs is not in itself unusual, and we believe that this habit contributed to the location and ingestion of toxic doses of cyanobacterial neurotoxin. Why dogs should be attracted to toxic benthic cyanobacterial mats or crusts is unknown. Cyanobacteria, and other aquatic and terrestrial microbes such as actinomycetes, produce many taste and odour compounds which have hitherto been of interest due to their adverse impact on aquaculture and drinking water quality⁶. Perhaps these are attrac-

tive to opportunistic omnivores such as the domestic dog. The significance of cyanobacterial taste and odour compounds as (fatal) attractants to animals has yet to be investigated.

G. A. Codd

C. Edwards

K. A. Beattie

Department of Biological Sciences,
University of Dundee,
Dundee DD1 4HN, UK

W. M. Barr

Duisdale,
Lochgilphhead,
Argyll PA31 8NJ, UK

G. J. Gunn

Scottish Agricultural College,
Veterinary Investigation Centre,
Inverness IV2 4JZ, UK

Dating of Nazca aqueducts

SIR — It is true that, to the best of our present knowledge, Spanish colonial documents do not settle the question of the ultimate origin of the underground aqueducts (*puquios*) of Nazca¹. However, at least one document does refer "specifically to the Nazca galleries". This is housed in Peru's national archive, and is part of the records of the viceregal water court². It establishes and confirms the water distribution of the Visambra *puquio* from 1692 to 1778, incorporating papers from the 1730s. In 1698 several haciendas, a rectory and the Indian community of Nazca had to give up flow for two hours every morning so that water could pass through the town of Nazca.

What has been dated by atomic mass spectroscopy (AMS) is not necessarily the construction of the *puquios*, or the time the two stones sampled were placed in the canals, but rather the times when the stones were last trimmed by masons. In prehispanic, colonial and republican Andean architecture, the reuse of ancient dressed stone is very common. New vertical segments of the Cantalloq *puquio* have been constructed in living memory³ and it is worrying that Dorn *et al.* do not pinpoint the locations where their samples were taken⁴.

In 1954, archaeologists observed that the Nazca to Inca site of Tambo Viejo in the nearby Acari valley was being disassembled and its stones trucked away for use in an unidentified irrigation project⁵. Water flowed through this site, so the reuse of stones from Tambo Viejo would not be incompatible with the observation reported, but not illustrated, by Dorn *et al.*, that the morphology of *puquio* varnish indicates that it formed in moist conditions. This is only circumstantial evidence, but it is enough to

caution us that, as Bray writes¹, the matter of *puquio* dating is only "partially resolved".

As yet, a complete and accurate map of the Nazca *puquios* has not been published. Furthermore, large portions of the *puquios* consist of trenches dug a few metres to the water table, roofed with wood or stone lintels supporting earthen or rubble fill to ground level. It is these stone lintels that produced the AMS dates. To the best of my knowledge, no *puquio* has ever been excavated. A careful and extensive examination of fill contents might reveal much about the sequencing of *puquio* construction, a process that could not have been accomplished in a single moment. It is strange that archaeologists have not applied their most basic and characteristic methodology to the problems presented by the underground aqueducts of Nazca.

Monica Barnes

377 Rector Place,
Apartment 11J,
New York, New York 10280, USA

1. Bray, W. *Nature* **358**, 19 (1992).
2. Archivo General de la Nación [Peru], Sección Colonial, Juzgado de Aguas 3.3.7.23 (1778) (cited in Barnes, M. & Fleming, D. *Latin American Antiquity* **2**, 48–68; 1991).
3. Schreiber, K. Lancho Rojas, J. *Boletín de Lima* **59**, 51–62 (1988).
4. Dorn, R. I., Clarkson, P. B., Nobbs, M. F., Loendorf, L. L. & Whitley, D. S. *Ann. Ass. Am. Geogr.* **82**, 136–151 (1992).
5. Menzel, D. & Riddell, F. A. *Archaeological Investigations at Tambo Viejo Acari Valley, Peru 1954* (California Institute for Peruvian Studies, 1986).

Geomagnetic reversal paths

SIR — Valet *et al.*¹ concluded that there is no statistical evidence for a preferred sector of longitude in the distribution of mean virtual geomagnetic pole (VGP) paths during reversals. This contradicts the earlier suggestion that the transitional field VGPs have a tendency to follow the longitudinal band over the Americas and to a lesser extent its antipode², an idea which we have used to infer aspects of core-mantle dynamics³. We believe that the statistical analysis of Valet *et al.* has led them to an incorrect conclusion.

The criteria used by the authors of both studies^{1,3} to select reliable records of reversals are basically the same: publication in refereed journals, existence of a reasonable number of transitional directions (usually 4 or 5), and sufficient description of the laboratory methods so that at least an idea of the magnetic cleaning is obtained. Thus there are only limited differences in the database of the two groups. These concern mainly the acceptance² or rejection¹ of four records from Argentina now published in a refereed journal⁴ and, more important, of

1. Francis, G. *Nature* **18**, 11–12 (1878).

2. Carmichael, W. W. in *Natural Toxins: Characterization, Pharmacology and Therapeutics* (eds Ownby, C. & Odell, G. V.) 3–16 (Pergamon, Oxford, 1989).

3. Codd, G. A. & Beattie, K. A. *PHLS Microbiol. Digest* **8**, 82–86 (1991).

4. Edler, L., Ferno, S., Lind, M. G., Lundberg, R. & Nilsson, P. O. *Ophelia* **24**, 103–109 (1985).

5. Gunn, G. J. *et al. Vet. Rec.* **130**, 301–302 (1992).

6. Persson, P.-E., Whitfield, F. B. & Krasner, S. W. (eds) *Wat. Sci. Technol.* **25**(2), 1–343 (1991).

records of the Blake, Cobb and Gilsa event. Despite these differences, the two databases are overall rather similar in terms of the distribution paths. For instance, Valet *et al.* also find that 43% of the records lie over the Americas and 20% in the antipodal band, and this makes a net 63% in the two longitudinal bands, which is almost exactly the number given by Tric *et al.* The different conclusions of the two groups thus arise from different analysis of the data.

Tric *et al.*² interpreted their results based on the visual impression given by rose diagrams, a standard technique of displaying circular distribution in which each record has equal weight. Although this did not introduce any bias towards the most detailed records, a rigorous statistical treatment is clearly desirable.

Valet *et al.*¹ first considered each record individually, and obtained a mean VGP longitude with a statistical weighting of the record. Then they divided the data into longitudinal bins and used the χ^2 test, with 18 sectors of 20° width, to assess their conclusions. It is, however, well known that the χ^2 test can be used only if the expected frequencies in each bin are at least equal to a minimum, which most authors set to five. The approximation improves with larger values, but significant errors are expected below this limit. If k denotes the number of groups, n must be at least $5k$. Some authors even suggest $10k$. In the calculation of Valet *et al.* the expected frequencies in each sector are $30/18 = 1.67$ (for the interval 0–2 million years), and $44/18 = 2.44$ for the 0–12-Myr interval, well below the acceptable limit for the χ^2 test to be valid. With 30 data, the maximum number of groups should be six, leading to 60° sectors, with 44 data it should be 8.8, leading to eight 45°, or nine 40° sectors. Interestingly, Valet *et al.* find χ^2 values rejecting the random hypothesis when 40 and 60° sectors are used.

Finally, when using the χ^2 test in binned circular statistics, the results must not depend on a particular (and possibly subjective) choice of the group interval. With the data of Valet *et al.*, if a series of eight 45° sectors is displaced in steps of 5°, so that after nine successive displacements the initial configuration is retrieved, different values of χ^2 are obtained for each of the sector orientations, four of which reject the random hypothesis while five do not. Clearly, the binned χ^2 statistics loses information and therefore requires many more observations than there are available to reach a robust conclusion.

After discussion with Professor S. Rao of the University of California, Santa Barbara, we now think that it is more appropriate to use the methods of circular statistics which, although somewhat

unfamiliar to palaeomagnetists, are widely used in biology (see, for example, ref. 5). These methods do not require any *a priori* partitioning of the data into longitudinal bins, thus avoiding subjective choices of the width and position of intervals on the circle. Furthermore, they allow one to judge whether a given nonrandom circulation distribution is unimodal or bimodal. This is particularly relevant because we want to test the hypothesis that there are two antipodal preferred bands of longitudes.

We have applied three of the most usual tests of circular statistics, the Rayleigh, Rao and Watson's tests, to the database of Valet *et al.* When the data are tested for unimodality, the results of the three tests are slightly contrasting, especially when only the 0–2 Myr period is considered. In contrast, when the data are tested for bimodality the results are clear: all three tests reject the hypothesis of random distribution at the 1% level for the 0–12 Myr period, and the Rayleigh and Watson tests also reject it for the 0–2 Myr period. Thus, when all the available data are used, circular statistics applied to the data set of Valet *et al.* support the hypothesis that there are two preferred bands of longitude over the Americas and its antipode.

The main problem is clearly the paucity of good transitional records. It may well be that future results will upset our

conclusions. However, with the data available today, the suggestion of the two preferential bands of longitude for transitional VGPs appears as a reasonable hypothesis which requires a physical explanation, either as a sedimentary artefact, as has been suggested⁶, or as a true geomagnetic phenomenon, as we believe. This latter view seems to be supported by recent results based on volcanic rocks^{7,8}.

Carlo Laj

Alain Mazaud

Robin Weeks

Centre des Faibles Radioactivités,

Laboratoire Mixte CEA-CNRS,

91198 Gif-sur-Yvette, France

Michael Fuller

Department of Geological Sciences,

University of California Santa Barbara,

California 93016, USA

Emilio Herrero-Bervera

Hawaii Institute of Geophysics,

Honolulu, Hawaii 96822, USA

1. Valet, J. P., Tucholka, P., Courtillot, V. & Meynadier, L. *Nature* **356**, 400–407 (1992).

2. Tric, E. *et al.* *Phys. Earth planet. Int.* **65**, 319–336 (1991).

3. Laj, C., Mazaud, A., Weeks, R., Fuller, M. & Herrero-Bervera, E. *Nature* **351**, 447 (1991).

4. Ruocco, M. *Geophys. Astrophys. Fluid Dyn.* **60**, 141–148 (1991).

5. Batschelet, E. *Circular Statistics in Biology* (Academic, 1981).

6. Langereis, C., van Hoof, A. A. M. & Rochette, P. *Nature* **358**, 226 (1992).

7. Hoffman, K. A. *Nature* **354**, 273–277 (1991).

8. Constable, C. *Nature* **358**, 230 (1992).

Questions in particle theory

SIR — The criticism by Kennedy¹ of Maddox's discussion² of an attempt to deal with the divergences of quantum field theories illustrates how entrenched present theoretical orthodoxy has become. It is certainly true that the renormalization procedures have dealt with the infinite integrals that appear in the quantum field theoretic algorithm. But to say that this *post hoc* procedure poses "no problem of principle" and that efforts to deal with this problem are "misguided and physically unnecessary" is simply to state the party line with which particle theorists are now trained.

The renormalization of electric charge in which factors of the form $e^2(1-K)$, where e is the so-called bare electron charge and K is a logarithmically divergent function, are simply defined to be equal to the observed value of the electron charge³ cannot fail to disturb the disinterested observer and is symptomatic of conceptual difficulties at the heart of present theory. We all know that quantum electrodynamics is a formidably accurate computational algorithm. But we all also know that quantum field theories are an uneasy way to satisfy the requirements of quantum mechanics and

the special theory of relativity, while the progeny of the liaison — the standard model and quantum chromodynamics — are forced contrivances with none of the elegance that comes from understanding at the very deepest levels.

It is no exaggeration to say that since the discovery of the renormalization procedures in the late 1940s theoretical particle physics has been a succession of gimmicks with which we have tried to outwit rather than understand nature. Shortly before his suicide in 1933, Ehrenfest⁴, in his paper 'A few queries related to quantum mechanics', foresaw the present trend: "These questions might well be cast aside as 'meaningless' for the sake of convenience. It is even considered proper to do so. But for this reason particularly, someone has to take upon himself the odium of asking these questions." So it is today.

Joseph Schwartz

2 Lancaster Drive,

London NW3 4HA, UK

1. Kennedy, D. C. *Nature* **357**, 449 (1992).

2. Maddox, J. *Nature* **353**, 497 (1991).

3. Bjorken, J. D. & Drell, S. D. *Relativistic Quantum Mechanics*, 158 (McGraw-Hill, New York, 1964).

4. Ehrenfest, P. *Z. Physik* **78**, 555 (1932).

Of squids and radar

M. F. Perutz

Chance and Design: Reminiscences of Science in Peace and War. By Alan Hodgkin. Cambridge University Press: 1992. Pp. 412. £40, \$59.95.

In August 1939, Alan Hodgkin and Andrew Huxley joined forces at the Laboratory of the Marine Biological Association in Plymouth for experiments on the transmission of nervous impulses along the giant axons of the squid. Hodgkin was a young research fellow at Trinity College in Cambridge and Huxley had just finished his undergraduate work in physiology there. On 23 August Hodgkin wrote to his mother: "We ran into a good many difficulties, but Andrew is a wizard at apparatus and got over them in an incredibly short space of time."

Huxley had tried to measure the viscosity of the axoplasm by letting mercury droplets fall through it, but had found to his surprise that it was a solid gel. This gave them the idea of introducing an electrode to measure directly the change in potential between the inside and the outside of the nerve during the propagation of an impulse. It was known already that the inside of a resting nerve fibre is electrically negative relative to the outside, but this negative potential was thought to fall to near zero during an impulse. Hodgkin and Huxley discovered instead that the amplitude of the potential change during the impulse was about double the resting potential, so that the internal potential became positive. These results opened the way to the discovery of the underlying physicochemical events, but just then

the Second World War broke out.

Hodgkin writes in his autobiography: "Only by undertaking some physically and intellectually demanding work could I forget the frustration engendered by having to abandon the marvellously exciting nerve experiments after 5 or 6 years' hard work, at the moment when they were most rewarding." He buried himself first in aviation medicine and then in the development of airborne centimetric radar for Bomber and Coastal Commands. Reconnaissance had shown that two-thirds of the bombs dropped at night over Germany, at heavy cost to crews and aircraft, fell in open fields. Radar helped the pilots to find their targets. These were not the railway yards and factories at which the Royal Air Force's communiqués pretended the bombs were directed, but cities, as they were all that radar could find. Hodgkin writes of his relief when the bombers were diverted to help with the invasion of France in the spring of 1944 — he "had grown utterly sick of the night bombing offensive".

Hodgkin reports that all scientists supported the view of Patrick Blackett and Henry Tizard that the development of airborne radar should be directed instead at the detection of German U-boats by Coastal Command, and he proudly relates that this eventually reduced allied shipping losses from the unsustainable figure of 600,000 tons a

month to only 50,000. He quotes Hitler's lament that "the temporary setback to our U-boats is due to a single invention of our enemies".

A temporary posting to Washington in 1944 led Hodgkin to visit the Radiation Laboratory at the Massachusetts Institute of Technology, where he was impressed by "the seemingly inexhaustible supply of people with a college degree in engineering" available to his colleagues working on radar; in Britain, he notes, "you were lucky to find anyone with a school certificate in physics". Ever since the war, this failing has been recognized as a major cause of the uncompetitiveness of much of British industry, but only now has that lesson sunk into technically illiterate governments, parliament and the civil service.

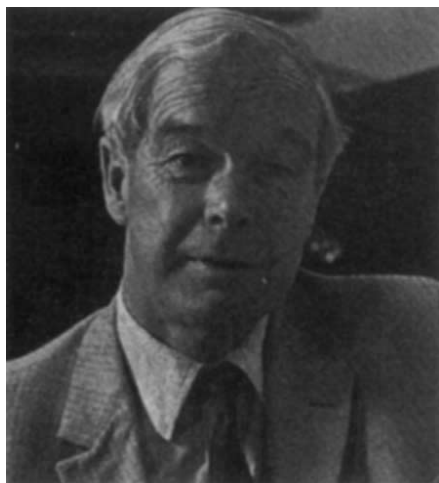
In June 1947, Hodgkin could at last return to Plymouth to resume his experiments on the giant nerve fibres of the squid. On 8 July, he wrote to his friend Victor Rothschild that he thought he had proved why the nerve membrane potential reverses during activity: "The reason is simply that the active membrane, instead of becoming freely permeable to all ions (as supposed by classic membrane theory) becomes much more permeable to Na than to K." In September 1947, Hodgkin and Bernard Katz proved that in the physiological range the reversed potential V_{Na} varied with the external sodium concentration as predicted by the Nernst equation

$$V_{Na} = \frac{RT}{F} \ln \frac{[Na] \text{ outside}}{[Na] \text{ inside}}$$

In the following year, Hodgkin and Huxley demonstrated rigorously and quantitatively that ionic currents across the nerve membrane are responsible for the propagation of nerve impulses. They



Goosander fishing: This picture was taken by R. and J. Kemp from behind a glass-fronted hide built into a river bank. It won the 1984 Wildlife Photographer of the Year competition (organized by BBC Wildlife Magazine and The Natural History Museum) and is featured, along with the winning and commended entries from 1991, in *Wildlife Photographer of the Year*. Published by Fountain (distributed in the US by Rockport) price £17.95, \$34.95.



Hodgkin in Cambridge, 1976.

measured the time courses of the changes in the permeability of the membrane to sodium and potassium ions when the membrane potential is suddenly changed, and showed that these changes accounted quantitatively for the amplitude and time course of the propagated impulse. To clinch their results, Hodgkin's graduate student Richard Keynes measured the ionic flux with radioactive sodium and potassium that he had prepared himself at the cyclotron in the Cavendish Laboratory at Cambridge. In 1952, Hodgkin and Huxley, both versatile at mathematics, developed differential equations to describe the propagation of the action potential in terms of the observed ionic currents.

It might be thought that any reasonably open-minded and perceptive scientist would have found this work convincing, but according to Hodgkin it met with widespread scepticism. Even the Nobel committee was unconvinced that nervous impulses are not transmitted by some energy-driven process along the axoplasm until 10 years later, when Peter Baker and Trevor Shaw replaced the axoplasm of a giant squid nerve with a salt solution of the same composition and showed that such a perfused nerve conducts nearly a million impulses without the intervention of any biochemical process. Scientists are apt to close their minds to new ideas, especially to those in their own fields.

There is much in this autobiography besides Hodgkin's classic work on neurophysiology. For the physicist, there is a technical account of the development of airborne radar. An introduction to the state of neurophysiology when Hodgkin started out in 1934, and to the physical principles of radar, would have been helpful to the uninitiated. For the general reader, the book contains vivid descriptions of Hodgkin's happy childhood among the often eccentric members of his many-branched Quaker family and of the Cambridge scene in

the 1930s. He tells us, for example, how as an undergraduate he joined a hunger march of the unemployed to London, accompanied by Guy Burgess, who wore a woollen jumper with a zip-fastener concealing an Eton tie, to be exposed in case he was arrested by the police. And when Anthony Blunt was elected a research fellow of Trinity College, the master, J. J. Thomson, famed for his discovery of the electron but apparently not for his tact, introduced him with the words: "This is the first time we have elected an art historian to a fellowship, and I very much hope it will also be the last."

Hodgkin's accounts of his first visit to New York as a Rockefeller Fellow and of a trip to Mexico in the late thirties are enlivened by the delightful letters written to his mother at the time. At the Rockefeller Institute he was agreeably surprised to be given a well-equipped laboratory, in contrast to Cambridge, where he had had to make all the equipment for his experiments himself.

Research and reality

F. E. G. Cox

Malaria: Obstacles and Opportunities, Edited by Stanley C. Oaks Jr, Violaine S. Mitchell, Greg W. Pearson and Charles C. J. Carpenter. *National Academy Press*: 1991. Pp. 309. \$48.

DESPITE all attempts at eradication, malaria not only continues to flourish but has re-emerged in areas from which it had been banished, has become increasingly resistant to drugs throughout its range and is being imported into nonmalarious countries on an unprecedented scale. In response to this growing threat, the US government, as one of the world's main financial contributors to research into and control of malaria, asked the Institute of Medicine to convene a multidisciplinary committee to assess the current status of malaria research and funding and to recommend strategies to tackle the problem. The results of the committee's deliberations are contained in this excellent volume. And in providing the background, the editors have in fact also produced a textbook on malaria, containing chapters on such topics as parasite biology, vector biology, epidemiology and vaccines. The book complements and updates Leonard Bruce-Chwatt's *Essential Malariology* (Heinemann, 1985) and should be read by everyone working on malaria.

Has the money spent on malaria been adequate, well spent and effective? The book reveals that between 1987 and 1989, the National Institute of Allergy

Hodgkin tells of his falling in love with Peyton Rous's eldest daughter Marni, and of their happy marriage after years of separation through the war, and finally of his receipt of the Nobel prize together with Huxley and John Eccles, who was awarded it for his fundamental work on synaptic transmission.

Hodgkin comes across in this enjoyable book as a genial and humane scientist passionately devoted to his research for its own sake both before and after being awarded his Nobel prize. For 36 years he did most of his experiments with his own hands and yet found time to have a family life, to develop wide interests in literature, painting and music, to watch birds and to cultivate lasting friendships with a great variety of fascinating people. There are no enemies. □

M. F. Perutz is at the MRC Laboratory of Molecular Biology, University Medical School, Hills Road, Cambridge CB2 2QH, UK.

and Infectious Diseases spent more than \$20 million on malaria, of which 61 per cent was devoted to vaccines and immunization and only 3.2 per cent to chemotherapy and 2.3 per cent to vectors. Figures such as these may help to explain the statements in the introduction that "the outlook for malaria is grim" and "in large part because of the spread of drug and insecticide resistance there are fewer tools to control malaria than there were 20 years ago". One can only speculate about the status of malaria had these figures been reversed. The committee's recommendations are even more interesting. In all, there are 26 of them. Unfortunately, some are predictable: "The committee recommends that increased funds be made available", "... recommends support for research training". But other recommendations, such as to screen new anti-malarial compounds and develop substances to restore the effectiveness of existing drugs, tackle the main problems. There is, however, nothing critical of the *status quo*, no suggestions for any shift in funding or radical changes in direction, and no novel solutions. Overall, there is too much emphasis on theory; for example, a lot of attention is given to epidemiological concepts, particularly stratification, which characterizes epidemiologic zones of malaria in terms of their main determinants, an approach that is bound to lead to a proliferation of the eight zones described and one which fudges the main differences between childhood malaria in Africa and adult malaria in Asia. Nevertheless, the

committee's recommendations do provide accurate summaries of the current approach to malaria, particularly in the prefaces to each chapter imaginatively headed "Where we want to be in the year 2010" and in the suggestions for research that are scattered throughout the book.

All but one of the chapters are anonymous, but probably the most interesting is the exception, a dissenting view by Awash Teklehaimanot from the World Health Organization in Geneva, who draws attention to the limitations of the epidemiological approach. He points to inadequacies of control and research in malarious countries and suggests that the logical approach is to rectify these deficiencies. This echoes many of the views put forward in G. A. T. Targett's edited volume *Malaria: Waiting for the Vaccine* (Wiley, 1991; for a review see *Nature* **356**, 300; 1992). The chapter puts the others in the book in context — overall the book is too research-dominated and lacks concrete ideas on how to control malaria in the immediate future.

The last words in the book are Teklehaimanot's: "I recommend that an in-depth review of malaria prevention and control be undertaken in order to come up with fundamental and realistic solutions." Many reading the book will agree — the review should put greater emphasis on what can actually be achieved, the redeployment of scarce resources, cost-effectiveness and, above all, the need to do something now, particularly for children who cannot wait for the year 2010. □

F. E. G. Cox is in the Division of Life Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

New in paperback

- **Too Hot to Handle: The Race for Cold Fusion** by Frank Close. Penguin, £6.99. Reviewed by Sir Brian Pippard in *Nature* **350**, 29 (1991), who said that it "should be read as an exemplary tale by all who are concerned about the conflicting demands of scientific integrity, personal ambition and public interest".
- **The Matter Myth: Beyond Chaos and Complexity** by Paul Davis and John Gribbin. Penguin, £6.99. For a review see *Nature* **356**, 632 (1991).
- **Technological Risk** by H. W. Lewis. Norton, \$11.95, £8.50. Reviewed in *Nature* **349**, 749 (1991).
- **The Vibrational Spectroscopy of Polymers** by D. I. Bower and W. F. Maddams. Cambridge University Press, £19.95, \$34.95.
- **The Fatigue of Materials** by S. Suresh. Cambridge University Press, £24.95, \$39.95.

Responsible development

Russell D. Thompson

Confronting Climate Change: Risks, Implications and Responses. Edited by Irving M. Mintzer. Cambridge University Press: 1992. Pp. 650. £50, \$80 (hbk); £19.95, \$34.95 (pbk).

THE social, economic, political and scientific implications of global climate change are set to dominate decision-making at all levels. This authoritative, up-to-date and superbly produced book examines the "risks, dilemmas and opportunities of the emerging political era" in a warming world, focusing on the possible dangers of climate change, the reliability of this knowledge and the best ways to respond to the changes.

The issues are tackled convincingly by 44 authors (scientists, economists and political analysts) from four continents, who were aided by 25 planners and 58 reviewers. The project was coordinated by the well-respected Stockholm Environment Institute and, indeed, the material represents an impressive continuation of the institute's long concern about inadvertent climate modification.

The text is organized into 23 chapters consisting of a review chapter on global warming and 22 others grouped into five sections. The first section on the science of climate change emphasizes the uncertainties and dilemmas associated with what is known about the climate system. A variety of chemical changes and linkages in the atmosphere are examined, particularly in terms of important feedback mechanisms. The implications of these processes for climate change are assessed through the interpretation of ice-core material and palaeoclimatic and historical evidence.

The second section on the impact of global climatic change assesses the environmental and geopolitical effects of global warming on future sea levels, food production, water resources, extreme weather events and disasters, and population dynamics. The third section on energy use and technology analyses the role of energy sources in the emission of greenhouse gases. The authors examine the environmental and socio-political consequences of uncontrolled use of fossil fuels and suggest technological alternatives and institutional responses for a low-emission future. Section four is devoted to economic issues and responses of international agencies, nongovernmental organizations and individual corporations. The economic benefits and social costs of reductions in greenhouse gas emissions are examined

in terms of two strategies which advocate either immediate action (the 'no regrets' principle) or caution and delay (the 'wait and see' policy).

In the final section, the authors address the problems of successfully implementing schemes and protocols to address global climate change, namely the equitable, global distribution of the costs and benefits. Particular attention is given to a workable, cooperative solution between developed and developing countries. There is also an urgent need for a stronger partnership between the public and private institutions dedicated to sustainable development (including the World Bank, the United Nations, the European Communities and the Organization of American States).

A useful glossary explains some of the scientific and economic jargon, and the editor's introduction provides a valuable synopsis of the text. Perhaps a better synthesis of the individual chapters could have been achieved by the inclusion of a final, overview chapter.

Nevertheless, with its level of commitment, expertise and specialization, this book cannot fail to become the leading statement on the consequences of climate change in the foreseeable future. □

Russell D. Thompson is in the Department of Geography, University of Reading, Whiteknights, Reading RG6 2AB, UK.

Post-Halley comets

David W. Hughes

Rendezvous in Space: The Science of Comets. By John C. Brandt and Robert D. Chapman. W. H. Freeman: 1992. Pp. 290. \$23.95, £16.95.

COMETARY science changed considerably after 1986. Before then, astronomers had seen about 700 comets. Roughly 550 of these had orbital periods of more than 200 years and had been glimpsed only once. The rest had short periods, some having been seen passing the Sun about 30 times. But our degree of ignorance about the physical, chemical, orbital and evolutionary states of each comet was very similar. Then in 1986 an armada of spacecraft flew past Comet Halley. Suddenly, we knew a lot about one member of the cometary family. For the first time, a low-resolution image of a comet's central nucleus was available; the composition of the dust and gas emitted by a comet had been crudely analysed by spacecraft instruments; and the interaction between cometary plasma and solar wind had been sampled.

Cometary scientists soon realized, however, that they had done little more than reveal the huge extent of their ignorance. Basic quantities such as the mass and density of Comet Halley were unknown, never mind the more interesting details of its origin, original size, internal structure and surface variability. And, worse still, astronomers had no idea whether Halley is typical of comets in general or is something of an odd-ball (Halley is relied upon for the calibration and interpretation of other cometary events).

The success of the 1986 European and Japanese cometary space missions seemed to put a nail in the coffin of NASA's cometary exploration programme. Not only had NASA missed the opportunity to rendezvous with Comet Temple-2 after visiting Comet Halley, but its plans to observe Halley from the Astro-1 space platform were quashed by the Challenger space-shuttle disaster. As a final curtain, the US Congress has now cancelled the Comet Rendezvous Asteroid Flyby spacecraft, which would have orbited a comet for a year or two before finally landing on its surface.

Brandt and Chapman seem to have two dilemmas. The publisher's blurb for *Rendezvous in Space* suggests that our understanding of comets has changed completely since the investigation of Comet Halley. The authors, quite rightly, will have none of this, and their book stands firmly on the shoulders of their excellent *Introduction to Comets* (Cambridge University Press, 1981). Strong emphasis is placed on how cometary scientists have been painstakingly revealing the secrets of comets for more than two centuries. Halley is just another chapter, albeit a rather long one.

The second dilemma concerns the book's level. Unfortunately, the authors provide a better introductory guide to cometary science in their 1981 book than in this newish work. (I use the word newish advisedly because the similarity between the two books is quite marked.) New science is often more difficult to understand than old science, so to appease the intended innumerate reader, Brandt and Chapman have written the first 211 pages without resource to any equations. All the relevant mathematics has been shuffled ignominiously into a set of appendices. Here one can revel in the joys of such things as plasma physics and the cometary surface energy balance. Detailed computer programs are given for calculating cometary ephemerides and there is a collection of 11 pictures aimed at providing an overview of the *in situ* measurements made of the plasma close to Comet Giacobini-Zinner and Comet Halley, and of the dust and gas near the latter. But the dichotomy is too marked; readers who

enjoy the appendices will find the preceding text too bland, whereas those who like the text will be swamped by the appendices.

This is a good book for newcomers, but old hands will be disappointed that these first-class authors have not doubled the length of their text and been more critical in their approach to such topics as the origin of comets, the prevalence of cometary impacts, the decay of comets and the characteristic range of members of the cometary family. I would have been happier if even more emphasis had been placed on the gaping holes of ignorance that pepper cometary science. □

David W. Hughes is in the Department of Physics, University of Sheffield, Sheffield S3 7RH, UK.

A selectionist view

Walter F. Eanes

The Causes of Molecular Evolution. By John H. Gillespie. Oxford University Press: 1992. Pp. 336. £25, \$35.

OVER the past decade, arguments for the neutral theory of molecular evolution have enjoyed prominence through books written by Motoo Kimura, Masatoshi Nei, Wen-Hsiung Li, Dan Graur and others. Now, John Gillespie, a population geneticist, provides a selectionist's view of molecular evolution. This book represents the fusion of his long-term interest in the evidence for unequal rates of molecular evolution at the amino-acid level and in mathematical models of selection in varying environments. What emerges is an assessment of the neutralist model of molecular evolution and an elaboration of the process by which he proposes selection occurs at the amino-acid level.

The book encompasses several related topics. First, Gillespie discusses his favourite case studies of so-called protein micro-adaptations, and summarizes what he feels are the generalities about electrophoretic variation that emerged from studies in the 1970s. Next, he examines the main studies of molecular evolution at the DNA level, the data and logic behind the molecular clock, and the neutralist contention that proteins evolve at a constant rate. He includes a useful discussion of the theory of point processes, which lays the foundation for his view that substitution rates are not constant, but reflect episodic selection.

About a third of the book is devoted

to mathematical models of selection in varying environments. This is a slippery topic because of the increased dimensionality, and opens a large number of problems for both the mathematician and the experimentalist. At this point, readers should be warned. This is not textbook population genetics, but diffusion theory. And it is not for the light-hearted — although discussed in a relaxed conversational style, the topic will be intractable to many. The general theory is important to Gillespie's overall theme, but presented in this much detail it becomes a distraction from the main issues. It will undoubtedly be skipped over by most readers. Nevertheless, the take-home lesson of Gillespie's models is that, in contradiction to earlier theory, selection in random environments can maintain genetic variation, and his mathematical demonstration of this capability is an important contribution. In the final chapters, Gillespie describes and assesses the neutral theory and its variants.

I like this book, and it is valuable to have many elements of the debate under one cover. I am sympathetic to Gillespie's case, although this is not to say that I agree with all his interpretations, or that he is without bias. For example, he is critical of the work on chemostat selection in *Escherichia coli*, which found no evidence for selection acting on electrophoretic polymorphisms, and he dismisses it because it does not live up to his experimental standards. Yet some other studies that are in my opinion of dubious value become his paradigm for selection. Gillespie's critique will not (nor should it) lead to the discarding of the neutral theory, which will continue to serve as a useful, albeit limited, null hypothesis. He nevertheless raises many important issues, such as the similarity of the patterns of allele frequency expected under both models, problems with assuming equilibrium populations and the importance of defining the spectrum of mutations.

The book is very much the gospel according to Gillespie, yet he makes many of his cases well, especially the argument for varying rates of substitution. As a proponent of the selectionist view he is not alone, and this work will begin to counter the influence that defenders of the neutral theory have enjoyed over the past decade. Given the explosion of molecular data that will emerge in the future, this is a timely contribution which will promote interest in the study of adaptation at the molecular level. □

Walter F. Eanes is in the Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794, USA.

Tectonic forcing of late Cenozoic climate

M. E. Raymo & W. F. Ruddiman

Global cooling in the Cenozoic, which led to the growth of large continental ice sheets in both hemispheres, may have been caused by the uplift of the Tibetan plateau and the positive feedbacks initiated by this event. In particular, tectonically driven increases in chemical weathering may have resulted in a decrease of atmospheric CO₂ concentration over the past 40 Myr.

THE past few years have seen increasing interest in the role that tectonic movements may play in controlling global climate, partly stemming from suggestions that the uplift of the Tibetan plateau in the late Cenozoic had a profound effect on atmospheric circulation^{1,2}. In addition, it has been proposed that tectonically driven increases in chemical weathering rates over the same interval caused a drawdown of atmospheric carbon dioxide, leading to global cooling^{3,4}. These hypotheses have provoked controversy in a wide range of fields. Geophysicists Molnar and England⁵ put forth the inverse proposal that climate changes were responsible for the inferred increases in elevation and erosion. Geochemists^{6–10} argued whether the ocean strontium isotope record could be used as a proxy for chemical weathering rates in the past. Climate modellers^{10–13} discussed whether chemical erosion could increase without a concomitant increase in mantle degassing of CO₂, for which there is no evidence.

All of these issues are interrelated, with the consequences of one interpretation affecting the others. Here we review the multidisciplinary aspects of the uplift-climate problem, in particular addressing those issues that have provoked the most discussion and controversy over the past few years. We start by presenting evidence for Cenozoic climate trends, and then discuss our uplift-erosion hypotheses, which we believe to be consistent with available late Cenozoic geological data.

Evidence for Cenozoic climate change

Because sedimentation occurs continuously in the deep sea, the best records of global Cenozoic climate variations come from the ocean. The oxygen isotopic composition of calcite, sensitive to both ice volume and ocean temperature, has been a particularly valuable tool for reconstructing past climate variations. As evaporation-precipitation processes preferentially concentrate the light isotope of oxygen, ¹⁶O, in continental ice sheets, a pronounced enrichment in the oceanic ¹⁸O/¹⁶O ratio occurs during periods of glaciation. A temperature-dependent ¹⁸O/¹⁶O fractionation between calcite and water also works in the same direction (colder temperatures cause ¹⁸O/¹⁶O enrichment); thus, the CaCO₃ shells of marine organisms recovered from deep-sea sediment cores isotopically record information about both ice volume and ocean temperature in the past.

Oxygen isotope data for the past 55 Myr (refs 14, 15) show a long-term increase in $\delta^{18}\text{O}$ values punctuated by times of relatively rapid change (Fig. 1). The sharp increase in isotopic values in the early Oligocene (~36 Myr BP) may reflect the first major ice-growth event on Antarctica¹⁶, which followed roughly 20 million years of cooling. Further increases in $\delta^{18}\text{O}$ in the middle Miocene and late Pliocene represent subsequent increases in ice volume in Antarctica and in the Northern Hemisphere respectively, although the exact portion attributable to ice volume rather than deep ocean cooling remains uncertain. Assuming no ice before the early Oligocene, Miller *et al.*¹⁴ infer deep-water temperatures up to 12 °C warmer than present in the early Eocene, and Shackleton and Kennett¹⁵ report data from the southern Pacific Ocean indicating deep and surface water temperatures up to 15 °C warmer in the early Eocene than at present.

In addition to the long-term cooling inferred from oxygen isotopes, enhanced differentiation of the mid-latitude continents into areas of wetter and drier climates is also indicated by the terrestrial fossil record (see ref. 17 for a summary of this evidence). At higher latitudes, Palaeocene subtropical flora have been identified in Alaska. Middle Oligocene vegetation at this latitude, however, indicates a temperature climate, and by the Pliocene, an essentially modern flora became established¹⁸. Likewise, progressive cooling of the polar regions is suggested by faunal macrofossil evidence from Ellsmere Island¹⁹. At high southern latitudes, the presence of Oligocene fossil pollen assemblages in Antarctica and Australia²⁰ suggests climates much warmer than present.

Overall, a preponderance of oceanic and terrestrial evidence points to a marked, progressive cooling of global climate beginning in the early Eocene. The most striking and convincing evidence of this climatic 'deterioration' is the abundant glacial detritus (tills, moraines, ice-rafted sediments and so on) which now blankets the mid-to-high latitudes of both hemispheres, both in the ocean and on land. Superimposed on this long-term cooling trend are shorter climate oscillations and steps, such as the rapid shifts in $\delta^{18}\text{O}$ at ~35 Myr and 15 Myr (Fig. 1). These inferred ice-growth events may be in response to abrupt changes in 'forcing', or they may reflect nonlinear (threshold) responses to more gradual changes in climate boundary conditions. In addition to the 'steps', short intervals of climate amelioration are observed in the $\delta^{18}\text{O}$ record in the early Oligocene, late Oligocene and Miocene. Wolfe¹⁸ inferred similar transient warming phases from floral macrofossil evidence from Alaska and the American northwest.

Climate forcing mechanisms

Since early last century, scientists have been searching for explanations for the climate cooling of the late Cenozoic^{21–23}. The drift of continents over the poles is a popular idea, particularly for the Palaeozoic (for example refs 24–26); but such drift towards the poles has been minimal over the past 100 Myr and cannot explain the observed magnitude of Cenozoic global cooling²⁷. Another family of climate-change mechanisms calls on tectonic movements that form sills or ocean gateways in climatically 'critical' locations. For instance, thermal isolation caused by the development of the circum-Antarctic current when Antarctica separated from South America and Australia has often been invoked to explain the rapid cooling and glaciation of Antarctica in the Cenozoic^{28,29}. These investigators suggest that greater meridional heat transport before the initiation of circum-Antarctic ocean circulation resulted in higher sea surface temperatures around Antarctica, effectively preventing continental ice sheets from developing. Experiments with general circulation models (GCMs)³⁰ suggest, however, that rather than decreasing the likelihood of glaciation, greater meridional heat transport would promote glaciation by increasing snowfall in an already cold polar region.

Likewise, the relationship between the emergence of the Isthmus of Panama and the initiation of Northern Hemisphere glaciation is unclear. The formation of the isthmus seems to pre-date widespread Northern Hemisphere glaciation by 0.5 to

2 Myr (ref. 31), and GCM simulations suggest that the deflection of warm Atlantic waters through the Panamanian passage would actually result in cooling in the high-latitude North Atlantic³². Thus, although local tectonic changes undoubtedly influence ocean circulation and heat transport, it is unclear how important these changes are in explaining either the stepwise coolings inferred from the Cenozoic $\delta^{18}\text{O}$ record or the long-term cooling trend observed for the Tertiary.

Many early 'ice age' theories focused on possible effects that uplift and erosion of mountain ranges and plateaux could have on climate. Here we review new variations on these old themes; in particular, we propose that uplift of the Tibetan plateau has been the main driving force behind Cenozoic climate change. Although uplift during the Cenozoic has been inferred for many regions, including the Alps, East Africa, and parts of the Cordillera of North and South America (see ref. 17 for a summary), we focus on the Tibetan plateau because there is unequivocal evidence that it has been created during the Cenozoic, and because its unique size and location make it central to our hypotheses.

With a mean elevation of almost five kilometres and an area half the size of the United States, the Tibetan plateau is the most imposing topographic feature on the Earth's surface. It formed as a result of the collision of the Indo-Australian plate with the Asian plate; hard collision of these continents probably began in earnest in the middle Eocene (44–52 Myr ago; refs 33–35) and continues today. Little information is available on the elevation history of the plateau. In fact, we have only two firm data points: the modern elevation, and the presence of Cretaceous to lower Eocene marine limestones which show that much of the plateau was below sea level around 70–100 Myr ago with some areas still under water until ~52 Myr (refs 33–35). Although we know that the plateau got from point "a" to point "b", so to speak, a more detailed knowledge of its elevation history will clearly be essential both for understanding plate tectonic and geophysical processes^{33–35} and for evaluating hypotheses that propose a link between climate change and plateau elevation.

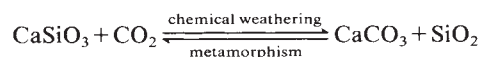
One such hypothesis^{1,2,36} notes that the modern Tibetan plateau is so high and so broad that it not only drives a regionally intense monsoon circulation, but also perturbs atmospheric circulation on a hemispheric scale. Ruddiman and colleagues^{1,2,36} propose that the absence of this plateau before the Eocene would have resulted in significant climatological differences around the Northern Hemisphere. To test this idea, Ruddiman and Kutzbach² initiated a series of experiments using the National Center for Atmospheric Research (NCAR) GCM of the Earth's atmosphere in which they changed only the topography of the present landmasses. Their results (see also ref. 36) showed that many of the global changes in precipitation and temperature inferred by geologists and palaeobotanists to have occurred over the past 40 Myr were consistent with differences in the climates calculated by the model. For instance, relative to the zero-topography experiment, the GCM run with modern topography shows a decrease in summer precipitation around the Mediterranean caused by increased subsidence of dry air masses and stronger northeasterly winds. Similarly, the spread of summer-dry vegetation and increased dustiness in the Mediterranean region since the Miocene has been inferred from geobotanical and sedimentological evidence.

The palaeobotanical record reflects a trend from equable, moist temperate climates in the early and mid-Cenozoic to cold/warm and dry/wet regional patterns in the Northern Hemisphere today. The climate differences in the GCM experiments with and without topography are consistent with this trend toward increased seasonality and regional differentiation of climate. But as discussed by Ruddiman and Kutzbach², the GCM experiments do not show a pronounced drop in high-latitude summer temperatures as elevations are increased, suggesting that development of high topography alone is not

sufficient to initiate the growth of large terrestrial ice sheets in both hemispheres. Experiments using GCM models incorporating important feedback processes (snow albedo, sea ice, mixed-layer ocean temperatures) produce coolings that are larger by almost half than earlier versions (J. Kutzbach, personal communication). For the winter season, many regions cool by amounts that represent a considerable portion of the temperature changes inferred from geological records. For the crucial summer ablation season which controls ice volume, however, the amount of cooling remains far short of that required to explain long-term Cenozoic cooling. Additional factors, such as changes in atmospheric composition, seem to be required to explain global Cenozoic cooling.

The Cenozoic 'icehouse effect'

It is widely accepted that changes in the concentration of radiatively important trace gases in the atmosphere (particularly CO_2) can alter the Earth's climate. A direct link between atmospheric temperature and CO_2 concentration is observed throughout the 140,000-year Vostok ice-core record of the last glacial-interglacial cycle³⁷, and many investigators have proposed a link between the evolution of global climate over the Cenozoic and changes in the composition of the Earth's atmosphere (for example refs 3, 23, 38–41). On timescales longer than a million years, atmospheric CO_2 levels are primarily controlled by the balance between the rate of volcanic input from the Earth's interior, and the rate of output through chemical weathering of rocks at the Earth's surface. These reactions can be simplified as



Two types of climate hypotheses have emerged: those in which variations in the atmospheric CO_2 levels are driven primarily by changes in volcanic outgassing, the main input term^{38,40–42},

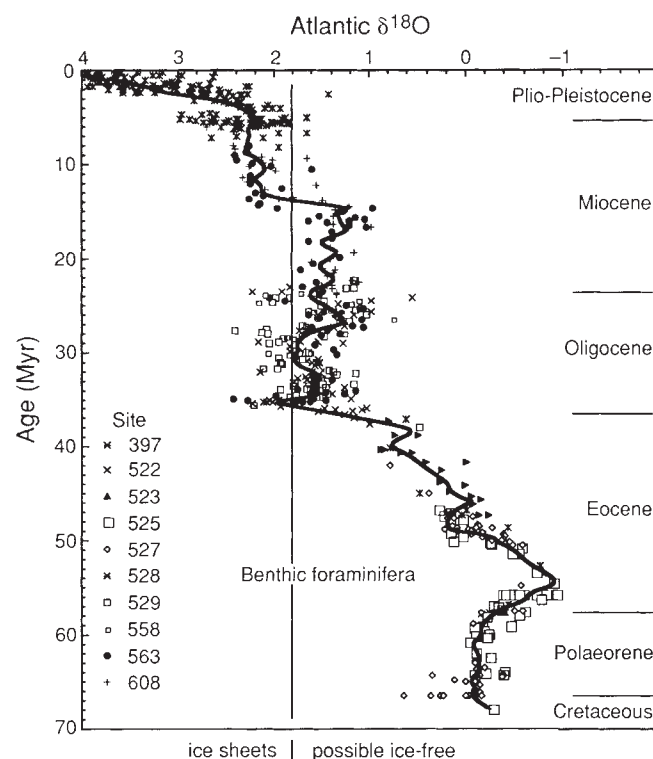


FIG. 1 Compilation of benthic $\delta^{18}\text{O}$ measurements from Deep Sea Drilling Program sites spanning the past 70 Myr. The long-term increase in $\delta^{18}\text{O}$ values reflects cooling of the deep ocean and growth of ice sheets at high latitudes (after ref. 14). $\delta^{18}\text{O} = [({}^{18}\text{O}/{}^{16}\text{O})_{\text{sample}}/({}^{18}\text{O}/{}^{16}\text{O})_{\text{standard}} - 1]$, where standard is PDB.

and those in which CO₂ variations are driven by changes in chemical erosion rates, the main output term^{3,4,23}. The first class of models predicts an association between times of high CO₂ levels and high rates of chemical weathering; in contrast, the erosion models predict an association between low CO₂ levels and high chemical weathering rates. We evaluate these two predictions later.

Berner, Lasaga and Garrels³⁸ (hereafter referred to as BLAG) developed a numerical model of CO₂ inputs (through metamorphism and decarbonation reactions) and outputs (through chemical weathering) to the atmosphere over the past 100 Myr. They related the flux of CO₂ into the atmosphere linearly to the rate of seafloor production and hence subduction, and the output flux of CO₂ from the atmosphere to the area of continents available for chemical weathering, with atmospheric temperatures acting as a strong negative feedback (assuming that higher temperatures cause more chemical weathering). As formulated, their model produces higher CO₂ levels in the Cretaceous because of (1) more rapid seafloor spreading and volcanic outgassing of CO₂ at that time, and (2) a decrease in the land area available to chemical erosion, because of globally higher sea levels. The higher sea levels are in large part related to faster seafloor spreading rates and increased ridge volume.

Although the BLAG model and subsequent modifications^{41,42} correctly yield higher global temperatures in the Cretaceous, the timing of the principal temperature change indicated by these models does not match the timing of the major climate cooling inferred from the geological record. In the BLAG models, the largest decline in atmospheric CO₂ and global temperatures occurs between 100 and 50 Myr ago (pre-Eocene), driven primarily by decreases in global seafloor spreading rates which control mantle outgassing and, indirectly, sea level. But geological evidence⁴³ suggests that the greater climate cooling occurred primarily between 50 Myr and the present (post-Eocene).

A possible explanation for this mismatch is suggested by erosion-driven CO₂ models^{3,4,23}. On the basis of modern river studies^{44,45}, Raymo *et al.*³ suggested that the long-term removal rate of CO₂ from the atmosphere by chemical erosion is a strong function of continental relief rather than area, and that atmospheric temperature exerts a relatively weak control on global chemical erosion rates. The logic is as follows: today, much of the Earth's land surface is characterized by low-lying, deeply weathered shield areas which contribute little to the global river flux of solutes. Even though chemical weathering is pervasive in these regions, the absolute rates are low. In contrast, mountainous regions are dominated by rapid mechanical erosion which, in turn, increases the surface area of fresh minerals available for chemical attack. Three further factors enhance chemical breakdown of detrital grains in mountainous regions: the abundance of easily weathered sedimentary silicates derived from uplifted passive margin sequences (weathering of carbonates does not ultimately alter atmospheric CO₂), the orographic concentration of rainfall on mountain slopes, and steep slopes which flush away mechanical and chemical erosion products, constantly exposing fresh minerals.

These effects are most marked in the Himalaya. The collision of India with Asia resulted not only in the uplift, deformation and erosion of an antecedent passive margin, but also in the formation of the Tibetan plateau where incident solar heating in summer drives the strong atmospheric convection and rainfall associated with the Asian monsoon. (Numerical models suggest that the high topography of the Indian subcontinent is largely responsible for the intensity of the Asian monsoon^{2,36,46-48}.) Data from the eight largest rivers draining the Himalayan-Tibetan region (Table 1) show that almost 25% of the total dissolved load reaching the ocean today comes from a watershed area that represents only 5% of the Earth's land surface^{49,50}. Thus, because of geologically unusual circumstances (the pres-

TABLE 1 Total river dissolved fluxes from Himalayan-Tibetan region

River	Dissolved solids (10 ⁶ tons yr ⁻¹)	Drainage area (10 ⁶ km ²)
Ganges/Brahmaputra	323.5	1.669
Yangtze (Chiang)	166	1.827
Si Kiang	132	0.464
Irrawaddy	91	0.414
Indus	62	0.916
Mekong	60	0.849
Huangho	33.5	0.814
Percentage of global total	25%	4.2%

Fluxes are from refs 49, 50.

ence of a large plateau near a warm ocean), a disproportionate fraction of the Earth's chemical weathering occurs in this small region in Asia. The monsoon precipitation that we invoke is not the 'normal' orographic rainfall found in areas where prevailing westerlies or easterlies impinge on meridional mountain ranges, but is caused by a thermally driven northward flux of moisture from the equatorial Indian Ocean. The Tibetan plateau lies in the latitude of the westerlies, but far from the main Atlantic moisture source, and 'normal' orographic precipitation (on the western slopes) is small.

On the basis of these and similar observations, we propose that late Cenozoic uplift of the Himalayan region and Tibetan plateau would have resulted in regionally, and hence globally, higher chemical erosion rates, causing a drawdown of atmospheric CO₂ and global cooling. The timing of this tectonically driven CO₂ decrease should be post-Eocene, coincident with the formation of the Tibetan plateau and in agreement with geological evidence for when global cooling was most rapid.

Proxies for chemical weathering rates

One important difference between the two types of hypotheses described above is the prediction of when chemical weathering rates are greatest: in BLAG-type models, weathering rates increase as global temperatures rise, whereas in our model, weathering rate increases are correlated with falling global temperatures. To test these predictions, we need a proxy for global chemical erosion rates. Such a proxy may be the ⁸⁷Sr/⁸⁶Sr isotope record preserved in marine carbonates^{3,6-9,51-55}.

The seawater ⁸⁷Sr/⁸⁶Sr ratio recorded by marine carbonates (Fig. 2a) reflects a balance between input of radiogenic material, relatively high in ⁸⁷Sr/⁸⁶Sr, weathered from continents (average value in present-day runoff is 0.7119; ref. 51) and nonradiogenic material of low ⁸⁷Sr/⁸⁶Sr introduced by hydrothermal activity (average value 0.7035). The isotopic composition of the river strontium flux is not invariant but depends on the lithology of the rocks being weathered⁵¹⁻⁵³. Typically, carbonate weathering results in high river fluxes of strontium with low isotope ratios (<0.709), which would tend to buffer the oceanic value, whereas silicate weathering releases a low flux of relatively radiogenic strontium (>0.710; refs 8, 51, 53). An exception to this relationship has been proposed by Palmer and Edmond⁸ who speculate that metamorphic repartitioning of elements in the orogenic zone of the Himalayas has resulted in the formation of extremely radiogenic (>0.720) easily weathered carbonates. Measurements of the isotopic composition of Himalayan limestones are needed to confirm this hypothesis. A third strontium input to the ocean, from redissolution of marine carbonates (average value 0.7084; ref. 51), buffers the isotope ratio of the oceanic strontium reservoir towards the integrated mean value of the preceding 10-15 Myr.

The pronounced post-Eocene rise in ⁸⁷Sr/⁸⁶Sr ratios (Fig. 2a) could be due to an increase in the delivery of radiogenic strontium from land, or to a decrease in seafloor hydrothermal activity. The widely held assumption that hydrothermal activity

is a function of seafloor spreading rates, which are estimated from marine magnetic anomalies, can be used to estimate the hydrothermal flux of nonradiogenic strontium through time. Seafloor spreading rates seem to have changed little over the past 30–40 Myr (ref. 56), and it is widely agreed that the pronounced late Cenozoic increase in the oceanic $^{87}\text{Sr}/^{86}\text{Sr}$ ratio can therefore only be explained by a considerable increase in the delivery of radiogenic strontium from land. This could have been caused by increases in global chemical weathering rates^{3,6,9,12} and/or the mean river isotopic composition of strontium^{8–10}.

Given the association between silicate weathering and high $^{87}\text{Sr}/^{86}\text{Sr}$ values, there are only two ways that rivers could become more radiogenic over the past 40 Myr without increasing silicate weathering rates. One is to decrease dramatically the delivery of low $^{87}\text{Sr}/^{86}\text{Sr}$ ratio strontium from carbonate weathering, requiring an unlikely reduction in the weathering rates of these types of rocks. The second, mentioned above, is to repartition radiogenic strontium by metamorphism, from resistant granites and gneisses into easily weathered carbonates⁸. Palmer and Edmond⁸ attribute the unusually high strontium isotopic composition of the Ganges–Bramaputra river system to such ‘orogenic’ repartitioning. But a related study⁵³ of the Ganges–Bramaputra headwaters proposes that the high $^{87}\text{Sr}/^{86}\text{Sr}$ ratios in these rivers are in fact due to intense chemical weathering of granites and gneisses (silicates).

Many studies of the Cenozoic strontium record^{3,6,7,9,12,53} have concluded that the pronounced Cenozoic increase in oceanic $^{87}\text{Sr}/^{86}\text{Sr}$ ratios was due to increased rates of chemical weathering over this interval. Richter *et al.*⁹ suggest that the strontium flux to the ocean has increased by ~40% over the past 40 Myr, an estimate which conservatively assumes that the strontium isotopic composition of Himalayan rivers increased from a ‘normal’ global value of 0.7105 in the Oligocene to its present anomalously high value of 0.7127, and that hydrothermal exchange was proportional to the rate of seafloor generation. In addition, this study and others^{6,53} concur with the suggestion³ that erosion in the Himalaya has been responsible for a significant fraction of the observed increase in $^{87}\text{Sr}/^{86}\text{Sr}$. But the fraction attributable to erosion in low-latitude mountainous regions such as the Himalayas, as opposed to erosion by high-latitude continental ice sheets or alpine glaciers elsewhere, is poorly constrained. By assuming a link to more easily assessed rates of mechanical erosion, Hodell *et al.*⁶ estimate that up to 50% of the increase in oceanic $^{87}\text{Sr}/^{86}\text{Sr}$ values over the past five million years may have been due to chemical erosion of the Canadian shield by the Laurentide ice sheet. Likewise, Miller *et al.*⁵⁷ proposed that glacial erosion in Antarctica was responsible for much of the increase in $^{87}\text{Sr}/^{86}\text{Sr}$ since the Eocene.

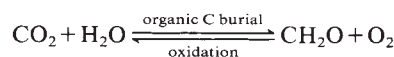
Although the evolution of the seawater strontium curve is consistent with a pronounced increase in chemical weathering since the Eocene, this interpretation is clearly not unique; a considerable fraction of the $^{87}\text{Sr}/^{86}\text{Sr}$ change could be due to variations in the isotope ratios in source rocks. The interpretation is further complicated by the fact that some strontium (albeit at generally low isotopic values) derives from the weathering of carbonates, a process which does not result in any net removal of CO_2 from the atmosphere. Ideally, we need a proxy for just the silicate component of weathering. Only with a number of weathering proxies, possibly including carbonate accumulation rates^{3,58,59} or Ge/Si ratios⁶⁰, will we be able accurately to define global weathering rates in the geological past.

The need for negative feedbacks

The erosion hypothesis described above predicts that chemical weathering rates can increase without a concomitant increase in mantle degassing rates. Increasing the erosional drawdown of CO_2 while holding the volcanic input constant would, however, deplete the atmosphere of all its CO_2 within a few million years¹⁰. By invoking tectonics as the main control on

weathering intensity, we implicitly negate the importance of the temperature–weathering feedback which prevents a runaway ‘icehouse’ (or ‘greenhouse’) in BLAG-type carbon cycle models. Hence, a different negative feedback is needed, one that adds CO_2 back to the ocean–atmosphere reservoir without adding alkalinity (for instance, Ca^{2+}) which removes CO_2 . Three processes that could provide such a feedback are discussed: imbalances in the organic carbon subcycle, precipitation of silicate minerals in the deep sea and seafloor weathering of basalt.

In addition to the carbonate–silicate cycle, the organic carbon subcycle, described schematically as



has a small but significant impact on the cycling of atmospheric CO_2 . On average over the Phanerozoic, roughly 80% of the carbon removed from the ocean–atmosphere system was deposited as carbonates, whereas ~20% was buried as organic matter⁶¹. Because organic matter is extremely enriched in ^{12}C during photosynthesis, any change in the relative size of this reservoir through time would be reflected in a change in the $^{13}\text{C}/^{12}\text{C}$ ratio of the world’s ocean and atmosphere (the surface carbon reservoir from which the organic matter is derived, or added to). If, for instance, organic carbon burial decreased over the past 40 Myr, the ocean–atmosphere reservoirs would become enriched in ^{12}C , the $^{13}\text{C}/^{12}\text{C}$ ratio of marine calcite would decrease, and a net addition of CO_2 to the atmosphere would occur. Such a change in the relative fraction of organic carbon buried would affect atmospheric CO_2 levels independently of global silicate weathering and mantle degassing rates.

The most complete compilation of bulk carbonate $\delta^{13}\text{C}$ values during the Cenozoic (Fig. 2b; ref. 61) shows just such a decrease. Using these data, Shackleton⁶¹ estimated that the fraction of carbon buried as organic matter rather than carbonate dropped from 20% to 10% over the course of the Neogene, causing a net addition of CO_2 to the atmosphere (alternatively, the data are consistent with an increase in kerogen weathering relative to organic carbon burial, also leading to a net addition of CO_2 to the atmosphere). Such a change in the organic carbon subcycle is of the correct magnitude and direction to prevent a tectonically sustained increase in silicate weathering (of ~40%) from completely stripping the atmosphere of CO_2 . Estimates depend on assumptions of how photosynthetic fractionation factors changed, when to define ‘steady state’ (or the present long-term mean) and what carbon isotope data set is used. Similar studies using different assumptions and data bases have come to more conservative conclusions^{12,41,42}.

Why would the long-term burial rate of organic carbon have decreased over the past 40 Myr? Most organic carbon burial today takes place in shallow marine environments such as estuaries, continental shelves and deltas. In these well oxygenated environments, the burial rate of organic carbon is a positive function of the total sediment burial rate⁶². But because carbon isotope data suggest that less organic carbon was being buried in the late Cenozoic, while other evidence suggests that worldwide sedimentation rates have been increasing^{63–65}, it seems that this was not the main factor controlling organic carbon burial in recent geological history. On the other hand, increased erosion of continental shelves, estuaries and deltas as sea level dropped through the Cenozoic could have resulted in a decrease in organic carbon burial in the shallow marine realm⁶¹ and an increase in the erosion of previously deposited organic-rich sediments.

A second possible control on organic carbon burial is nutrient supply: a decrease in organic carbon burial may have been caused by a decline in the amount of phosphorus and nitrogen available to form organic matter. This does not seem to be the case for the past 40 Myr, as the nutrient supply from rock weathering has, if anything, increased (as discussed above), and

because much dissolved phosphorus currently entering the ocean is removed by inorganic precipitation⁶⁶.

A third possible control is the oxygen concentration in the deep ocean. Today many areas of the deep ocean are nearly anoxic⁶⁷. Even lower dissolved oxygen concentrations could result in a considerable increase in marine organic carbon burial^{68–70}, although this point is controversial^{71,72}. The oxygen content of the deep ocean is controlled by the temperature of the water as it leaves the surface, with O_2 concentrations increasing as sea surface temperatures decrease. Today deep waters form in polar regions with a dissolved O_2 concentration of $\sim 350 \mu\text{mol kg}^{-1}$. If high-latitude sea surface temperatures were $\sim 10^\circ\text{C}$ warmer in the mid-Cenozoic, a level consistent with planktonic isotopes^{14,73}, then O_2 concentrations would have been only $\sim 275 \mu\text{mol kg}^{-1}$. The inferred increase through time in ocean oxygen concentrations could have resulted in decreased organic carbon burial.

Deep-ocean oxygen concentrations would decrease and sea level would fall as CO_2 concentrations in the atmosphere began to fall: as the climate gets colder, oxygen solubility in sea water increases and continental ice growth causes sea level to drop. Each of these changes, in turn, could lead to decreased organic carbon burial and addition of CO_2 to the atmosphere, providing a negative feedback to climate change. Increased continentality and aridity, associated with cooling global temperatures, could also have resulted in the shrinkage and drainage of freshwater marshes and swamps, again leading to a decrease in the burial of organic-rich sediments.

As originally pointed out by Shackleton⁶¹, the bulk carbonate $\delta^{13}\text{C}$ data provide convincing evidence that there has been a net addition of light carbon (and CO_2) to the ocean and atmosphere since the Oligocene. We propose that this carbon flux

would have partially counteracted a drawdown of atmospheric CO_2 by enhanced erosion of silicates. Before invoking organic carbon burial as an effective negative feedback to the global carbon cycle, however, one also has to consider the relative residence times of CO_2 (<1 Myr) and O_2 (~ 10 Myr) in the atmosphere. Any decrease in organic matter burial will result in net removal of O_2 from the atmosphere. Eventually the continued drawdown of O_2 will cause the partial pressure of oxygen in the atmosphere to fall, probably leading to organic carbon burial rates increasing again. Thus, the negative feedback to the carbonate-silicate cycle provided by the organic carbon subcycle would gradually become less effective five to ten million years after the 'event' that initiated the change, a problem which may not be critical in the case discussed here; namely, gradually more intense forcing as the Tibetan plateau increases in elevation.

Two other processes have been proposed as negative feedbacks to atmospheric CO_2 decreases driven by enhanced silicate weathering. One is precipitation of silicate minerals on the sea floor (ref. 67, and W. Broecker, personal communication). Although this process is poorly understood (see summaries in refs 67, 74), much of the sodium and potassium entering the sea must be precipitated as silicate minerals. On the other hand, calcium derived from silicate weathering is mostly precipitated as carbonate, resulting in the net removal of CO_2 from the atmosphere. As atmospheric CO_2 falls, however, ocean pH rises (OH^- rises), creating conditions more favourable to the precipitation of Ca as deep-sea silicates rather than carbonates. In this way, one could again decouple silicate weathering rates from atmospheric CO_2 levels. Seafloor weathering of basalt, described in the carbon-cycle model of François and Walker¹², provides a third mechanism to decouple silicate weathering, alkalinity and atmospheric CO_2 levels. In this process, hydrogen ions in ocean water react with seafloor basalts, releasing calcium ions which then combine with ocean carbon to form calcium carbonate. The net result is the removal of CO_2 from the ocean-atmosphere reservoir. But as atmospheric CO_2 falls, ocean pH and OH^- rise, creating conditions less favourable to the removal of CO_2 by this mechanism, again providing a negative feedback to a runaway 'icehouse'.

Given the suggestions above, it is certainly possible, if not probable, that increases in silicate weathering rates could be sustained for millions to tens of millions of years without an increase in mantle degassing rates. In particular, the assumption that rates of volcanism and silicate weathering need to be balanced on timescales longer than a million years¹⁰ may not be necessary. One last possibility is that seafloor spreading rates are not a good proxy for mantle degassing rates. Caldeira¹³ suggests that increased subduction of pelagic carbonates since the Jurassic has resulted in an increased metamorphic CO_2 flux from the mantle which is responsible for the higher rates of chemical erosion inferred for the late Cenozoic. This could provide the 'extra' CO_2 needed to sustain high silicate weathering rates at a time when seafloor spreading rates are constant.

The role of positive feedbacks

Molnar and England⁵ proposed an interaction between climate and tectonics that on first reading seems the exact inverse of our uplift hypothesis, but in fact may provide a fully complementary source of positive feedback. They argue that late Cenozoic global cooling was responsible for increased mechanical erosion, which through isostasy would create mountain ranges with high peaks and deep valleys, rather than uplift causing climate changes and erosion. They suggested two mechanisms by which climate change intensified erosion: expansion of mountain glaciers around the world; and changes in atmospheric circulation, and thus in precipitation patterns. At the limit, this hypothesis might be interpreted as suggesting that high mountain ranges existed throughout the Cenozoic, and that the only change during the Cenozoic has been the effect of climate change

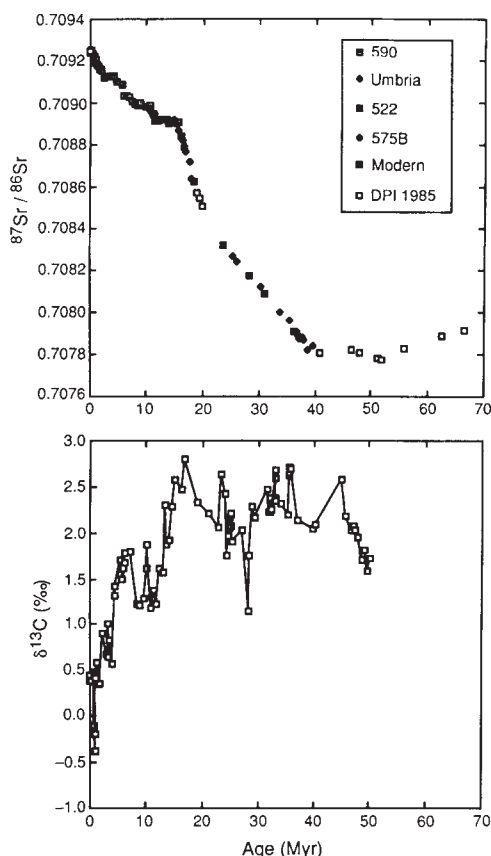


FIG. 2 a, Strontium isotopic composition of sea water for the past 70 Myr, based on analysis of marine carbonates (from ref. 9). b, $\delta^{13}\text{C}$ of bulk carbonate over the Cenozoic (from ref. 61). $\delta^{13}\text{C} = [(^{13}\text{C}/^{12}\text{C})_{\text{sample}} / (^{13}\text{C}/^{12}\text{C})_{\text{standard}} - 1]$, where standard is PDB.

on mountain relief. Such an interpretation would, however, leave the cause of Cenozoic cooling and circulation changes unresolved. In fact, Molnar and England⁵ noted the need for a first cause of Cenozoic cooling, and acknowledged Tibetan uplift as a possible cause of climate change. Their basic disagreement is with the inference of accelerating late-Cenozoic uplift throughout the world.

It is consistent with both hypotheses to argue that uplift of Tibet (and possibly elsewhere) is a plausible first cause of Cenozoic climate changes (through circulation changes and weathering), but that climate change (in particular, glacier activity) then caused additional erosion, exhumation and isostatic uplift in other regions, including some mountain belts regarded as orogenically 'dead' from the standpoint of horizontal plate tectonic motions. Thus, the arguments of Molnar and England can be viewed as a positive feedback mechanism, whereby uplift-induced erosion initiates global cooling, which then causes further glacial erosion and cooling worldwide.

An important implication of the plateau uplift hypothesis is that even steady-state plate motions can lead to non-steady-state effects on climate and, hence, possibly global relief. Continent-continent collisions which result in plateaux of the magnitude of Tibet are infrequent and episodic. Consequently, the accompanying effects of uplift (perturbations of atmospheric flow, large monsoons, and intense erosion and weathering), should be comparably infrequent. Over the past 700 Myr, only two other time periods were characterized by Tibetan-size plateaux, the late Precambrian and the late Palaeozoic⁷⁵; these were also intervals of widespread continental glaciation⁴. In summary, despite the continuous presence throughout geological history of high mountain terrain along the convergent margins of the world, it may be the rarer occurrence of plateaux that can drive climate away from steady state and decouple rates of horizontal and vertical tectonic movement.

Evaluating the effects on climate

We propose that over the past 40 Myr, uplift of the Tibetan plateau has resulted in stronger deflections of the atmospheric jet stream, more intense monsoonal circulation, increased rainfall on the front slopes of Himalayas, greater rates of chemical weathering and, ultimately, lower atmospheric CO₂ concentrations. These changes in climate may initiate strong positive feedbacks to global cooling through glacier-driven erosion⁵. A negative feedback, acting either through the organic carbon subcycle or elsewhere, must also occur to prevent the atmosphere from being completely stripped of CO₂. This model implies that atmospheric CO₂ levels are not maintained at steady state by the temperature-weathering feedback during times of intense tectonism (which may last many millions of years).

To evaluate the linkages between uplift, weathering and climate proposed here, proxy data for global chemical weathering rates over the Cenozoic and a better understanding of the elevation history of the Tibetan plateau are needed. We also need a quantitative understanding of the long-term carbon cycle, in particular the magnitude and source of variations in the organic subcycle and seafloor weathering. Ultimately, direct determination of atmospheric CO₂ levels over the past 100 Myr will be essential if we are to evaluate the uplift-erosion model^{3,4} and other carbon-cycle climate models^{38,41}. The most promising technique for this is the inference of atmospheric CO₂ levels from measurements of $\delta^{13}\text{C}$ in marine organic matter⁷⁶⁻⁷⁸. □

M. E. Raymo is at the Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. W. F. Ruddiman is at the Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia 22903, USA.

- Ruddiman, W. F. & Raymo, M. E. *The Past Three Million Years: Evolution of Climatic Variability in the North Atlantic Region* (eds Shackleton, N. J., West, R. G. & Bowen, D. Q.) 227-234 (Cambridge Univ. Press, Cambridge, 1988).
- Ruddiman, W. F. & Kutzbach, J. E. *J. geophys. Res.* **94**, 18409-18427 (1989).
- Raymo, M. E., Ruddiman, W. F. & Froelich, P. N. *Geology* **16**, 649-653 (1988).
- Raymo, M. E. *Geology* **19**, 344-347 (1991).
- Molnar, P. & England, P. *Nature* **346**, 29-34 (1990).
- Hodell, D. A., Mead, G. A. & Mueller, P. A. *Chem. Geol. (Isotop. Geosci. Sec.)* **80**, 291-307 (1990).
- Capo, R. C. & DePaolo, D. J. *Science* **249**, 51-55 (1990).
- Palmer, M. R. & Edmond, J. M. *Geochim. cosmochim. Acta* **56**, 2099-2111 (1992).
- Richter, F. M., Rowley, D. B. & DePaolo, D. J. *Earth planet. Sci. Lett.* **109**, 11-23 (1992).
- Berner, R. A. & Rye, D. M. *Am. J. Sci.* **292**, 136-148 (1992).
- Sundquist, E. T. *Quat. Sci. Rev.* **10**, 283-296 (1991).
- François, L. M. & Walker, J. C. G. *Am. J. Sci.* **292**, 81-135 (1992).
- Caldeira, K. *Nature* **357**, 578-581 (1992).
- Miller, K. G., Fairbanks, R. G. & Mountain, G. S. *Paleoceanography* **2**, 1-19 (1987).
- Shackleton, N. J. & Kennett, J. P. *Init. Rep. Deep Sea Drilling Proj.* **29** (ed. Kennett, J. P.) 801-807 (U.S. Government Printing Office, Washington DC, 1975).
- Barrett, P. J., Elston, D. P., Harwood, D. M., McKelvey, B. C. & Webb, P. N. *Geology* **15**, 634-637 (1987).
- Ruddiman, W. F., Prell, W. L. & Raymo, M. E. *J. geophys. Res.* **94**, 18379-18391 (1989).
- Wolfe, J. A. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **9**, 25-57 (1971).
- McKenna, M. C. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **30**, 349-362 (1980).
- Kemp, E. M. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **24**, 169-208 (1978).
- Lyell, C. *Principles of geology* (Murray, London, 1875).
- Dana, J. D. *Am. J. Sci.* **22**, 305-334 (1856).
- Chamberlin, T. C. *J. Geol.* **7**, 545-584; 667-685; 751-787 (1899).
- Crowell, J. C. & Frakes, L. A. *Am. J. Sci.* **268**, 193-224 (1970).
- Caputo, M. V. & Crowell, J. C. *Geol. Soc. Am. Bull.* **96**, 1020-1036 (1985).
- Crowley, T. J., Mengel, J. G. & Short, D. A. *Nature* **329**, 803-807 (1987).
- Barron, E. J. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **50**, 45-61 (1985).
- Kennett, J. P. *J. geophys. Res.* **82**, 3843-3860 (1977); *Mesozoic and Cenozoic Oceans* (ed. Hsu, K. J.) 119-122 (Am. geophys. Un., Washington DC, 1986).
- Barker, P. F. & Burrell, J. *Antarctic Geoscience* (ed. Craddock, C.) 377-385 (Univ. of Wisconsin, Madison, 1982).
- Oglesby, R. J. *Clim. Dynam.* **3**, 135-156 (1989).
- Keigwin, L. D. *Science* **217**, 350-353 (1982).
- Maier-Reimer, E., Mikolajewicz, U. & Crowley, T. *Paleoceanography* **5**, 349-366 (1990).
- Mercier, J.-L., Armijo, R., Tapponnier, P., Carey-Gailhardis, E. & Lin, H. T. *Tectonics* **6**, 275-304 (1987).
- Molnar, P. *Phil. Trans. R. Soc. Lond. A* **326**, 33-88 (1988); *Am. Scientist* **77**, 350-360 (1989).
- Harrison, T. M., Copeland, P., Kidd, W. S. F. & Yin, A. *Science* **255**, 1663-1670 (1992).
- Kutzbach, J. E., Guetter, P. J., Ruddiman, W. F. & Prell, W. L. *J. geophys. Res.* **94**, 18393-18407 (1989).
- Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408-414 (1987).
- Berner, R. A., Lasaga, A. C. & Garrels, R. M. *Am. J. Sci.* **283**, 641-683 (1983).
- Barron, E. J. & Washington, W. M. *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 546-553 (Am. geophys. Un., Washington DC, 1985).
- Walker, J. C. G., Hays, P. B. & Kasting, J. F. *J. geophys. Res.* **86**, 9976-9782 (1981).
- Lasaga, A. C., Berner, R. A. & Garrels, R. M. *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 397-410 (Am. geophys. Un., Washington DC, 1985).
- Berner, R. A. *Science* **249**, 1382-1386 (1990); *Am. J. Sci.* **291**, 339-376 (1991).
- Crowley, T. J. & North, G. N. *Paleoclimatology*, 339 (Oxford Univ. Press, New York, 1991).
- Gibbs, R. J. *Science* **156**, 1734-1737 (1967).
- Stallard, R. F. & Edmond, J. M. *J. geophys. Res.* **88**, 9671-9688 (1983); Stallard, R. F. *Physical and Chemical Weathering in Geochemical Cycles* (eds Lerman, A. & Meybeck, M.) 225-246 (Kluwer, Dordrecht, Holland, 1988).
- Manabe, S. & Terpstra, T. B. *J. Atmos. Sci.* **31**, 3-42 (1974).
- Hahn, D. G. & Manabe, S. *J. Atmos. Sci.* **32**, 1515-1541 (1975).
- Druyan, L. M. *J. Clim.* **2**, 127-139; 347-355 (1982).
- Meybeck, M. *Hydrol. Sci. Bull.* **21**, 265-284 (1976).
- Pinet, P. & Souriau, M. *Tectonics* **7**, 563-582 (1988).
- Palmer, M. R. & Edmond, J. M. *Earth planet. Sci. Lett.* **92**, 11-26 (1989).
- Palmer, M. R. & Edlerfield, H. *Nature* **314**, 526-528 (1985).
- Krishnaswami, S., Trivedi, J. R., Sarin, M. M., Ramesh, R. & Sharma, K. K. *Earth planet. Sci. Lett.* **109**, 243-253 (1992).
- Brass, G. W. *Geochim. cosmochim. Acta* **40**, 721-730 (1976).
- Burke, W. H. et al. *Geology* **10**, 516-519 (1982).
- Kominz, M. A. *Am. J. Petrol. Geol. Mem.* **46**, 109-127 (1984).
- Miller, K. G., Feigenson, M. D., Wright, J. D. & Clement, B. M. *Paleoceanography* **6**, 33-52 (1991).
- Davies, T. A., Hay, W. W., Southam, J. R. & Worsley, T. R. *Science* **197**, 53-55 (1977).
- Opdyke, B. N. & Wilkinson, B. H. *Paleoceanography* **3**, 685-703 (1988).
- Shemesh, A., Mortlock, R. A. & Froelich, P. N. *Paleoceanography* **4**, 221-234 (1989).
- Shackleton, N. J. *Marine Petroleum Source Rocks* (eds Brooks, J. & Fleet, A. J.) 423-434 (Geological Society of London, 1987).
- Berner, R. A. & Canfield, D. E. *Am. J. Sci.* **289**, 333-361 (1989).
- Donnelly, T. W. *Geology* **10**, 451-454 (1982).
- Hay, W. W. *Proc. int. Geol. Congr.* **27th** **6**, 15-38 (1984).
- Hay, W. W., Shaw, C. A. & Wold, C. N. *Geol. Rundsch.* **78**, 207-242 (1989).
- Froelich, P. N., Bender, M. L., Luedtke, N. A., Heath, G. R. & DeVries, T. *Am. J. Sci.* **282**, 474-511 (1982).
- Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* (Eldigio, Palisades, New York, 1982).
- Demailson, G. J. & Moore, G. T. *Bull. Am. Assoc. Petrol. Geol.* **64**, 1179-1209 (1980).
- Emerson, S., Fischer, K., Reimers, C. & Heggie, D. *Deep Sea Res.* **32**, 1-21 (1985).
- Emerson, S. *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 78-87 (Am. geophys. Un., Washington DC, 1985).
- Pedersen, T. F. & Calvert, S. E. *Am. Assoc. Petrol. Geol. Bull.* **74**, 454-466 (1990).
- Betts, J. N. & Holland, H. D. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **97**, 5-18 (1991).
- Keigwin, L. D. & Corliss, B. H. *Geol. Soc. Am. Bull.* **97**, 335-345 (1986).
- Drever, J. I. *The Geochemistry of Natural Waters* (Prentice-Hall, Englewood Cliffs, New Jersey, 1988).
- Dewey, J. F. & Burke, K. C. A. *J. Geol.* **81**, 683-692 (1973).
- Arthur, M. A., Dean, W. E. & Claypool, G. E. *Nature* **315**, 216-218 (1985).
- Jasper, J. P. & Hayes, J. M. *Nature* **347**, 462-464 (1990).
- Rau, G. H., Froelich, P. N., Takahashi, T. & DesMarais, D. J. *Paleoceanography* **6**, 335-347 (1991).

ACKNOWLEDGEMENTS. We thank W. Dietrich, R. Jeanloz, W. Broecker and P. Molnar for reviews. D. DePaolo and K. Miller provided copies of their figures. This work was supported by the NSF.

A model for the global variation in oceanic depth and heat flow with lithospheric age

Carol A. Stein* & Seth Stein†

* Department of Geological Sciences, University of Illinois at Chicago, Chicago, Illinois 60680, USA

† Department of Geological Sciences, Northwestern University, Evanston, Illinois 60208, USA

Variations in sea-floor depth and heat flow with age provide the main constraints on the thermal structure and evolution of the oceanic lithosphere. Joint fitting of heat flow and bathymetry yields a model with a hotter, thinner lithosphere than in previous models. The new model provides a significantly better fit to the data, including those from older lithosphere previously treated as anomalous. This will facilitate the analysis of lithospheric processes, including the effects of mid-plate volcanism and swells, regional subsidence, and hydrothermal circulation near spreading centres.

THE plate tectonic cycle, in which oceanic lithosphere cools as it spreads away from mid-ocean ridges and is reheated on returning to the mantle at subduction zones, is a surface manifestation of terrestrial convection¹⁻³ and the primary mode of heat transfer from the Earth's interior⁴⁻⁶. The resulting thermal evolution of the oceanic lithosphere governs its properties as a function of age and hence the style of plate tectonics.

Following the realization that sea-floor heat flow is highest at mid-ocean ridges and decreases with distance⁷⁻⁹, the systematic variation of ocean depth and heat flow with age (Fig. 1) became the primary constraint on models of the thermal evolution of the lithosphere. The two sets of data jointly reflect the evolution with age of the geotherm in the lithosphere, because the bathymetry depends on the temperature integrated over depth and the heat flow depends on the temperature gradient at the sea floor^{10,11}. The key features of the data, the decrease in heat flow and increase in sea-floor depth with age, prompted two classes of models describing how lithosphere cools as it spreads away from ridges. Each model predicts the variation in depth and heat flow with age. In one model¹⁰, the lithosphere behaves as the cold upper boundary layer of a cooling halfspace, such that the depth and heat flow vary as $\text{age}^{1/2}$ and $\text{age}^{-1/2}$, respectively. In the other model, the lithosphere is treated as a cooling plate with an isothermal lower boundary^{8,11}, and so behaves as a cooling boundary layer until it reaches ages at which the effects of the lower boundary cause the depth and heat flow curves to flatten and vary more slowly with age. The plate thickness in the model is the asymptotic thermal thickness approached by old lithosphere, which presumably reflects the combined effects of temperature and rheology¹. Parsons and Sclater¹² found that such a model, for a 125-km-thick plate with a basal temperature of 1,350 °C, provided good fits to the data then available. It describes the general shape of the depth curve, including the flattening for ages >70 Myr and the heat flow for ages >50 Myr. Its main limitation, the overprediction of heat flow for ages <50 Myr, is thought to result from the transport of significant amounts of heat by water circulation¹³, rather than by the conductive cooling assumed in such models.

The simplicity and ease of computation of the plate model makes it a useful representation for the global average of depth and heat flow with age, and for the thermal evolution of the lithosphere. The model's key feature, the isothermal base of the lithosphere, is a simple way of describing the additional heat assumed to prevent the halfspace cooling from continuing for older ages. Various mechanisms including radiogenic heat¹⁴, shear heating¹⁵, small-scale convection^{1,16} and mantle plumes¹⁷⁻¹⁹ have been proposed as sources of the additional heat. In most of these formulations, the asymptotic thickness to which the lithosphere evolves corresponds to the depth at which

the additional heat is supplied, and above which the temperature changes cause bathymetric variations. Because the thermal structure within the lithosphere does not depend critically on how this heat is supplied¹⁶, the plate model is a standard for comparing the predicted temperatures within the lithosphere to data including seismic wave velocities²⁰, flexure due to applied loads^{21,22} and the depths of intraplate earthquakes²³.

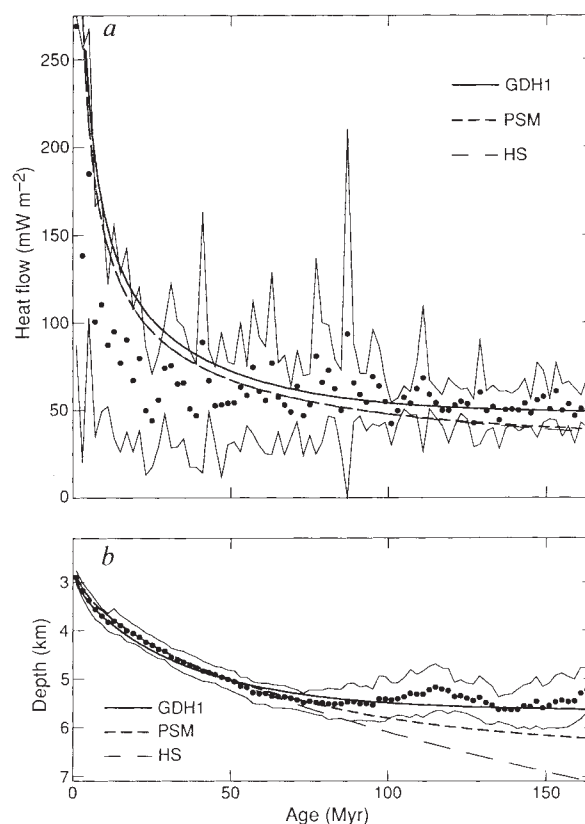


FIG. 1 Data and models for (a) heat flow and (b) ocean depth as a function of age. Depths are an average of North Pacific²⁴ (north of the Equator) and Northwest Atlantic²⁵ (15° N–45° N, 40° W–80° W) values; heat flow are from sites (Fig. 2) in these regions. The data are averaged in 2-Myr bins, and one standard deviation about the mean value for each is shown by the envelope. Shown are the plate model of Parsons and Sclater¹² (PSM), a cooling halfspace model with the same thermal parameters (HS) and the GDH1 plate model derived here. In a, the HS and PSM curves overlap for ages younger than ~120 Myr.

Much attention has been directed to depth^{24,25} and heat flow^{26,27} anomalies relative to the Parsons and Sclater model (a term used for a plate model with the parameters they estimated from the data), to identify processes not adequately described by the model. These studies have been impeded because this model, which we refer to as PSM, predicts deeper depths and lower heat flow than generally observed for lithosphere older than 70–100 Myr (Fig. 1). In particular, these biases pose difficulties for identifying possible mid-plate thermal perturbations due to mantle plumes^{28,29}.

To address these difficulties, we combine the depth and heat flow data to estimate suitable parameters for plate models. Largely by using improved heat-flow data, we find that a plate model with a hotter and thinner lithosphere (1,450 °C at the base of a 95-km-thick plate) satisfies many of the data that are anomalous with respect to PSM.

Data analysis

We used a heat-flow data set considerably larger than that available to Parsons and Sclater. Of the 5,539 data, 4,217 are from a published compilation³⁰, and the remainder, some of which have been reported elsewhere²⁹, are from one in preparation by D. Abbott and C.A.S. The data provide good coverage of the main ocean plates (Fig. 2). We used only good-quality experimental data, as defined in ref. 29, and excluded data from marginal basins, whose thermal evolution may differ from that of larger oceanic plates³¹. The data are at sites younger than 166 Myr, the oldest at which magnetic anomalies provide ages. Because the heat flow and depth for old lithosphere provide joint constraints, we derived the model using only data (Fig. 1) from the North Pacific and Northwest Atlantic, where the published depth data extend to at least 166 Myr and have been corrected for sediment thicknesses.

We initially compared the data to the heat flow and depth predicted by PSM and a halfspace model. In plate models^{11,12}, the temperature as a function of distance x from the ridge and depth z below the sea floor is given by

$$T(x, z) = T_m \left[z/a + \sum_{n=1}^{\infty} c_n \exp(-\beta_n x/a) \sin(n\pi z/a) \right]$$

where the asymptotic thermal plate thickness is a , its basal

temperature is T_m and

$$c_n = 2/(n\pi), \quad \beta_n = (R^2 + n^2\pi^2)^{1/2} - R, \quad R = va/(2\kappa)$$

As this expression is in terms of depth rather than height above the plate base, it differs slightly from that in ref. 12. The Peclet number R , relating the advective and conductive heat transfer, incorporates the half spreading rate v and the thermal diffusivity $\kappa = k/(\rho_m C_p)$, where k is thermal conductivity, ρ_m is mantle density and C_p is specific heat. The corresponding heat flow is

$$q(x) = k \left(\frac{\partial T}{\partial z} \right)_{z=0} = q_s \left[1 + 2 \sum_{n=1}^{\infty} \exp(-\beta_n x/a) \right]$$

where $q_s = kT_m/a$ is the asymptotic heat flow for old lithosphere, in which the thermal gradient is linear. The ocean depth varies with distance as

$$d(x) = d_r + d_s \left[1 - \frac{8}{\pi^2} \sum_{j=1}^{\infty} j^{-2} \exp(-\beta_j x/a) \right]$$

$$d_s = \frac{\alpha \rho_m T_m a}{2(\rho_m - \rho_w)}$$

where j takes only odd values, d_r is the depth of the ridge axis, d_s is the asymptotic subsidence for old lithosphere, α is the volume coefficient of thermal expansion and ρ_w is the water density. We refer ocean depths to the ridge axis, rather than to the less easily identified asymptotic depth for old lithosphere. For sufficiently young ages, a plate model is indistinguishable from a cooling halfspace model¹⁰, in which for age t the temperature is

$$T(t, z) = T_m \operatorname{erf}(z/(4\kappa t)^{1/2})$$

the heat flow is

$$q(t) = kT_m(\pi\kappa t)^{-1/2}$$

and the depth is given by

$$d(t) = d_r + 2(\kappa t/\pi)^{1/2} \alpha \rho_m T_m / (\rho_m - \rho_w)$$

The predictions shown (Fig. 1) are for the parameters used in PSM, a 125-km-thick plate with a basal temperature of 1,350 °C, coefficient of thermal expansion $3.28 \times 10^{-5} \text{ K}^{-1}$, specific heat $1.171 \text{ kJ kg}^{-1} \text{ K}^{-1}$ ($0.28 \text{ cal g}^{-1} \text{ K}^{-1}$), conductivity

FIG. 2 Locations and lithospheric ages of the heat-flow data. North Pacific and Northwest Atlantic data (Fig. 1) were used to derive the GDH1 model; all the data are used in Fig. 4a and b.

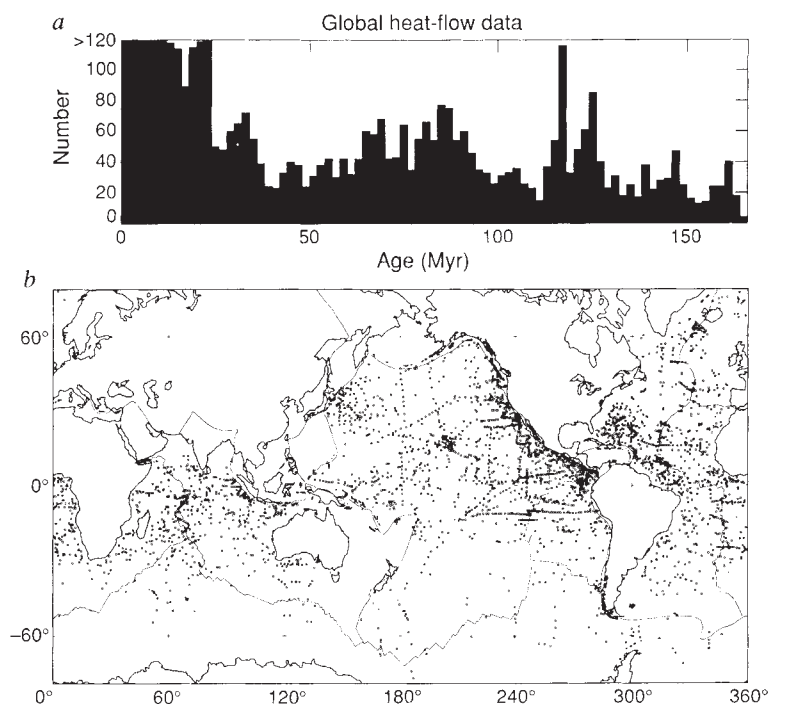
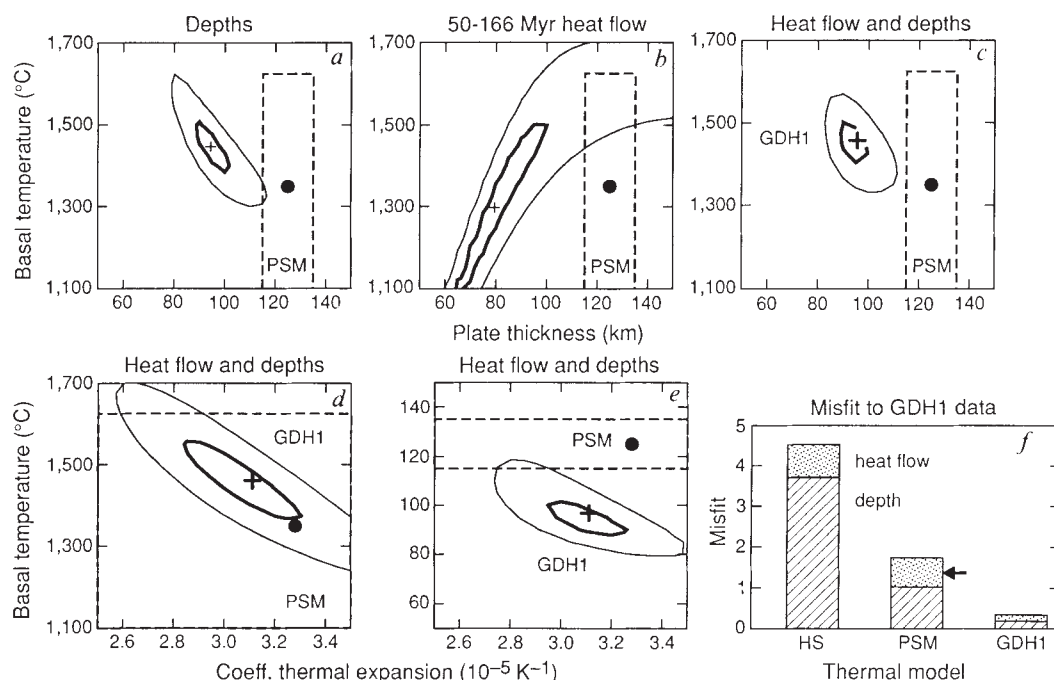


FIG. 3 *a–e*, Contour plots of the misfit to subsets of the data as a function of combinations of the three parameters. In each panel, the best fit is denoted by a cross, and the contours show values 1.25 and 2.5 times the minimum misfit. Parsons and Sclater's values (PSM) and their estimated uncertainties are also indicated. The best-fitting solution for the combined depth and heat flow (*c–e*), a 95-km-thick plate with a basal temperature of 1,450 °C and coefficient of thermal expansion $3.1 \times 10^{-5} \text{ K}^{-1}$, is model GDH1. *f*, Misfit to the data for GDH1, PSM, the halfspace model and a version of the PSM model including radioactivity (arrow), which predicts the same depths but higher heat flow than PSM.



$3.138 \text{ W m}^{-1} \text{ K}^{-1}$ ($7.5 \times 10^{-3} \text{ cal K}^{-1} \text{ cm}^{-1} \text{ s}^{-1}$), mantle density $3,330 \text{ kg m}^{-3}$, water density $1,000 \text{ kg m}^{-3}$ and ridge depth 2,500 m, and for a halfspace model with the same thermal parameters and ridge depth. The predicted depths for PSM flatten with respect to the halfspace model for ages $> 70 \text{ Myr}$, whereas the predicted heat flow differs from that for the halfspace only slightly for large ($> 140 \text{ Myr}$) ages, because the bottom boundary condition influences the integrated temperature at an age younger than that at which its effect reaches the surface.

It is useful to separate the data into four categories. The heat flow data can be divided into ages younger or older than $\sim 50\text{--}75 \text{ Myr}$. The data for younger ages are not fitted by the conductive models, presumably because of hydrothermal circulation¹³. The heat flow for older ages changes slowly with age, such that with our data the slope is not resolvably different from zero for ages greater than 80–90 Myr, and the average values systematically exceed those predicted by PSM^{26–29,32}. Similarly, the depth data behave differently for ages younger or older than 40–70 Myr. Younger values are reasonably well fitted by either PSM, other plate models, or halfspace models, whereas the older depths are generally overpredicted by PSM, as previously noted^{24,25,29,33–36}.

We investigated how well the data were fitted by plate models with different lithospheric thicknesses, basal temperatures and coefficients of thermal expansion. We considered the combined misfit to the depth and heat-flow data

$$s^2(a, T_m, \alpha) = s_d^2 + s_q^2 = \frac{1}{N} \sum_{i=1}^N (d_i - \hat{d}_i)^2 / \sigma_{d_i}^2 + \frac{1}{M} \sum_{j=1}^M (q_j - \hat{q}_j)^2 / \sigma_{q_j}^2$$

where d_i and $\hat{d}_i$ are the observed and predicted depths in the i th age interval, σ_{d_i} is the standard deviation of the depth for this interval, and q_j , $\hat{q}_j$ and σ_{q_j} are the corresponding heat flow quantities. We used the PSM values for the other parameters, except for a slightly deeper ridge depth of 2,600 m, in keeping with recent data^{34,36}. Our approach of treating the depth and heat flow as joint constraints is conceptually similar to Parsons and Sclater's, differing primarily in that the data are sufficiently densely sampled to permit numerical fitting.

We calculated the misfit at increments of 5 km in plate thickness, 25 K in basal temperature and $5 \times 10^{-7} \text{ K}^{-1}$ in coefficient of thermal expansion. Figure 3*a–c* shows contours of the misfit as a function of basal temperature and thickness, for the best-fitting coefficient of thermal expansion, $3.1 \times 10^{-5} \text{ K}^{-1}$. These plots show a slice of the three-dimensional model space with α fixed. Figure 3*d* shows the misfit as a function of the coefficient of thermal expansion and basal temperature, for the best-fitting thickness of 95 km, and Fig. 3*e* shows the misfit as a function of the coefficient of thermal expansion and thickness, for the best-fitting basal temperature, 1,450 °C.

The misfit contours for the depths show a diagonal pattern because the asymptotic depth is proportional to the product of plate thickness and basal temperature. The best fit is for a thinner plate and higher temperature than in PSM, consistent with the observed depths (Fig. 1) being shallower than predicted by PSM. The heat-flow data for lithosphere older than 50 Myr provide a separate constraint. The misfit contours are elongated diagonally in the opposite sense as for the depths, because the asymptotic heat flow depends on the ratio of basal temperature to thickness. As for the depths, the best fit occurs for a thinner plate and higher T_m/a than for PSM, because the observed heat flow exceeds that predicted by PSM. The coefficient of thermal expansion plots show the best fitting value, and the partial tradeoffs because the depths for young and old ages depend on the products αT_m and $\alpha T_m a$, respectively. Joint fitting of the combined depth and heat flow data gives a best fit for a 95-km-thick plate, with a basal temperature of 1,450 °C, coefficient of thermal expansion, $3.1 \times 10^{-5} \text{ K}^{-1}$, and all other parameters except the ridge depth the same as for PSM.

This model, which we call GDH1 (for global depth and heat flow), fits the data considerably better than PSM, especially for old lithosphere (Fig. 1). GDH1 predicts an asymptotic depth for old lithosphere of 5,651 m, considerably shallower than the 6,400 m for PSM. The asymptotic heat flow for GDH1 is 48 mW m^{-2} , roughly 40% higher than the value of 34 mW m^{-2} for PSM. As a result, the depth and heat flow misfits for GDH1 are $s_d^2 = 0.18$ and $s_q^2 = 0.16$, whereas the corresponding misfits for PSM are substantially higher, 1.01 and 0.70. An *F*-ratio test indicates that the improved fit is significant above the 99.9% level. Figure 3*f* shows that the improved fit of GDH1 relative to PSM is comparable to that for PSM relative to the halfspace model. An *F*-ratio test shows that GDH1 also provides an

improved fit above the 99.9% level relative to a version of the PSM model in which, in addition to the heat flow due to plate cooling, 4 mW m^{-2} is assumed to be generated by radioactivity¹², giving the same depths as PSM but a better fit to the heat flow ($s_q^2 = 0.34$).

A range of models gives fairly similar curves of depth and heat flow against age. For example, if we assume the PSM value for the coefficient of thermal expansion, we obtain a best-fitting thickness of 95 km and a basal temperature of 1400°C . Alternatively, varying both thermal conductivity and α yields a best fit for k of $3.25 \text{ W m}^{-1} \text{ K}^{-1}$, α of $3.1 \times 10^{-5} \text{ K}^{-1}$, a thickness of 95 km and T_m of $1,425^\circ\text{C}$. An F -ratio test shows that the improved fit from estimating the conductivity from the data is not significant even at the 90% level, because the resulting value is similar to that previously assumed. It is, however, gratifying that solving for the conductivity gives a value consistent with that usually assumed. Different choices of data also yield similar best-fitting parameters. For example, use of the heat-flow data for ages older than 75 Myr, rather than 50 Myr, yields a best fit for a thickness of 95 km, T_m of $1,475^\circ\text{C}$ and α of $3.05 \times 10^{-5} \text{ K}^{-1}$. This model is so similar to GDH1 that it predicts an asymptotic heat flow within 1 mW m^{-2} and an asymptotic subsidence within 3 m of those for GDH1.

GDH1 is thus representative of the plate models that fit the data well. Quantification of the true uncertainties in model

parameters is difficult, given the limitations of the data and model (discussed later). From examination of fits to the data, we estimate the plate thickness as $95 \pm 15 \text{ km}$, basal temperature as $1,450 \pm 250^\circ\text{C}$ and α as $3.1 \pm 0.8 \times 10^{-5} \text{ K}^{-1}$, where the uncertainties are one standard deviation, and do not include the uncertainty in the model or other parameters (such as conductivity). Alternatively, if we regard α as known *a priori*, we estimate the thickness and T_m uncertainties as 10 km and 100°C , respectively. Thus our solutions have plate thicknesses generally less than Parsons and Sclater's estimate ($125 \pm 10 \text{ km}$), basal temperatures comparable to their estimate ($1,350 \pm 275^\circ\text{C}$) and α comparable to their estimate ($3.2 \pm 1.1 \times 10^{-5} \text{ K}^{-1}$).

Applications of the model

Thermal models of the lithosphere are used in several ways for tectonic studies. First, they provide reference models to characterize the average variation in depth and heat flow with age for assumed 'normal' lithosphere. It is thus possible to identify regions that are anomalous with respect to a reference model, and to estimate how anomalous the depth and heat flow are. Second, a model predicts a geotherm for 'normal' lithosphere, and hence can be used to draw inferences about the processes giving rise to the temperature structure of both 'normal' and 'anomalous' lithosphere. These two applications are decoupled, in that it is useful to characterize 'normal' lithosphere and identify anomalies, although the assumed thermal structure presumably approximates a more complex situation.

GDH1 is a more useful global reference model than PSM, as it fits many of the data better for old lithosphere. It avoids the difficulty that most old lithosphere seems anomalous relative to PSM. The fact that many previously inferred heat flow and depth anomalies are successfully described by a single model bears out the importance of using both data types to constrain models.

The depth and heat-flow predictions for GDH1 are conveniently and accurately approximated using a halfspace model with the same parameters for young lithosphere, and the first term of the series solution with $R \gg \pi$ for older lithosphere¹². The depth (m) is related to the age (Myr) by

$$\begin{aligned} d(t) &= 2,600 + 365t^2 & t < 20 \text{ Myr} \\ &= d_r + d_s[1 - (8/\pi^2) \exp(-\kappa\pi^2 t/a^2)] & t \geq 20 \text{ Myr} \\ &= 5,651 - 2,473 \exp(-0.0278t) \end{aligned}$$

and the heat flow (mW m^{-2}) is

$$\begin{aligned} q(t) &= 510t^{-1/2} & t \leq 55 \text{ Myr} \\ &= q_s[1 + 2 \exp(-\kappa\pi^2 t/a^2)] & t > 55 \text{ Myr} \\ &= 48 + 96 \exp(-0.0278t) \end{aligned}$$

Hotspot swells. The GDH1 predictions are useful for investigation of the depth and heat-flow anomalies at swells associated with presumed hotspots. Two common end-member models for the swells are a thermal origin, in which the lithosphere is reheated at depth^{37,38}, or a dynamic origin, in which upwelling mantle causes uplift³⁹⁻⁴¹. The observation of heat flow along the Hawaiian swell higher than predicted for PSM was initially treated as consistent with reheating³⁸, because this model predicts large heat-flow anomalies. Subsequent studies showed that swell heat flow differs at most slightly from that for lithosphere of comparable ages²⁸, so that referring the data to PSM overestimated the anomalies. GDH1 fits the global heat-flow data (only part of which was used in deriving the model) significantly better than PSM (Fig. 4b) and is therefore a more appropriate reference. Thus although relative to PSM the anomaly is ~ 6 – 10 mW m^{-2} (depending on whether radioactivity is assumed), the anomaly relative to GDH1 is $\sim 4 \text{ mW m}^{-2}$. The heat-flow anomalies over the other swells are similarly reduced, thus favouring dynamic rather than reheating models, in accord with recent seismological results⁴².

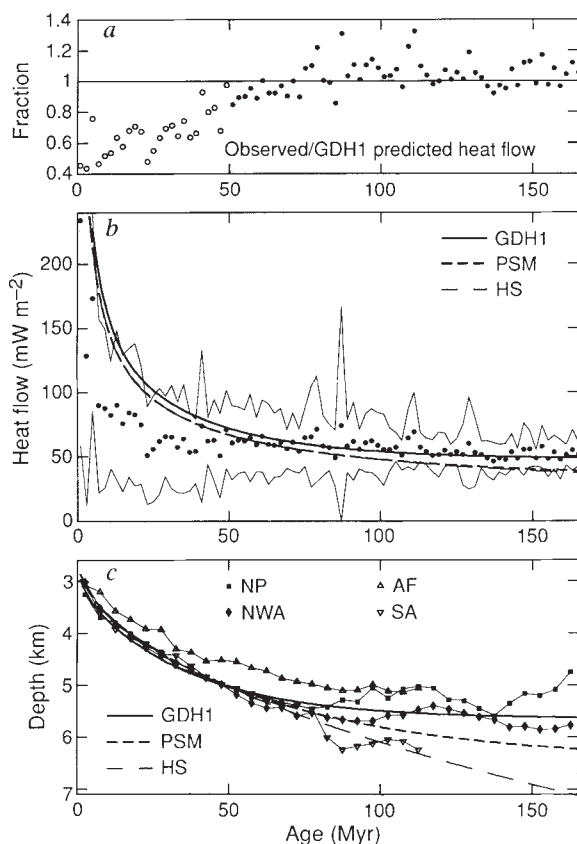


FIG. 4. Use of GDH1 as a reference model for depth and heat-flow analyses. *a*, Observed heat flow against age for the entire global dataset as a fraction of that predicted by GDH1. The discrepancy for ages younger than 50–75 Myr presumably indicates the fraction of the heat transported by hydrothermal flow. The data for ages < 50 Myr (open circles) were not used in deriving GDH1. *b*, Observed heat against age for the entire global dataset, and predictions of PSM, a halfspace model with the same thermal parameters (HS) and GDH1. The HS and PSM curves overlap for ages younger than ~ 120 Myr. Symbols are the same as in Fig. 1. *c*, Depth data (averaged in 5-Myr bins) for the South American (SA)⁴⁶ and African (AF)⁴⁶ plates, and the North Pacific (NP)²⁴ and Northwest Atlantic (NWA)²⁵ compared to the predictions of the three models. The differences between plates make any global reference model only an approximation.

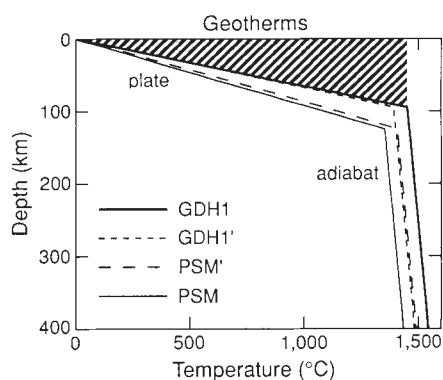


FIG. 5 Comparison of the asymptotic thermal structure for old lithosphere in models PSM and GDH1. GDH1, developed to fit higher heat flow, has a steeper geotherm and thus higher temperatures at depth. The sea-floor subsidence is proportional to the area (shaded) between the geotherm and $T = T_m$, so GDH1 predicts less subsidence and hence shallower bathymetry than PSM. Versions of these models incorporating the effect of radioactivity are labelled GDH1' and PSM'. Schematic adiabats are shown beneath the plate.

Hydrothermal circulation. A reference model like GDH1 can be used to study hydrothermal circulation in young lithosphere^{13,43,44}. Comparison of the observed conductive flux to that predicted (Fig. 4a) suggests that the fraction of heat transferred by fluid flow decreases steadily from less than ~50% in very young lithosphere. GDH1 slightly underpredicts the heat flow for old lithosphere, perhaps in part because it is also constrained by the depths (Fig. 3).

Depth anomalies. Although a reference model is needed to identify and characterize depth anomalies, the differences in subsidence rates between plates (Fig. 4c) and within plates are substantial enough^{10,34,45,46} that no model can fit all the data. Similarly, our fitting of mean depths, which are convenient for statistical analysis, or the alternative of fitting modal depths, which are less affected by processes that produce shallow depths²⁴, affects the result. Any global reference model is therefore an approximation to an ideal. Nonetheless, comparison with either mean or modal depths shows that GDH1 is a more useful reference than PSM, which has the unsatisfying implication that most older lithosphere is anomalously shallow. Moreover, the approach used to derive GDH1 may be applied to differences between and within plates. Although such analyses have often considered only depths and interpreted the differences in terms of temperature variations⁴⁶, it may sometimes be practical to use both data types and also test different plate thicknesses. Although the data for young lithosphere are insensitive to plate thickness and hence describable by halfspace cooling, the data for older lithosphere are sensitive to plate thickness. It would also be desirable to derive a revised global model, using sediment-corrected depth data for old lithosphere from other areas (such as the Indian Ocean and Northeast Atlantic).

Because GDH1 is derived by joint fitting of bathymetry and heat-flow data, both the depths and heat flow are assumed to reflect thermal structure. Regions anomalous with respect to the model can thus be characterized in terms of two end-member cases, thermally or topographically anomalous. The former situation (anomalously shallow bathymetry and high heat flow) presumably reflects temperatures in the lithosphere higher than in GDH1. In contrast, the latter situation (regions shallow with respect to GDH1, but with consistent heat flow) may reflect dynamic uplift, excess volcanism or thick crust. The small heat-flow anomaly associated with Hawaii and the other hotspot swells indicates that they are closer to this second case.

An important use of depth data is for analysis of the history of old lithosphere⁴⁷. The observation that old lithosphere is shallow with respect to a halfspace model gave rise to the plate model, which fits most of the data reasonably well and is thus the standard reference. The issue is generally not whether a given region of old lithosphere is anomalous relative to a cooling halfspace, because this is almost always the case, but whether it is shallow relative to old lithosphere of comparable age elsewhere²⁹. Despite the variations in the depth for a given age, such that 'flattening' is only an approximation, GDH1 describes most old lithosphere better than PSM. Thus the regions for ages

>100 Myr (primarily in the Pacific) which are anomalously shallow with respect to PSM (Figs 1, 4c) are either consistent with or considerably less shallow with respect to GDH1.

GDH1 or any reference model is thus helpful in investigating the physical process causing the flattening with respect to the square root of age curve. A reference model characterizes the average heat flow and depth values as a function of age, which models of the physical process must satisfy. Many recent studies favour hotspot reheating, in part because of questions about whether small-scale convection would cause flattening^{3,19}. If the flattening is caused by hotspots, therefore, a plate model like GDH1 describes their average effect, and regions shallow with respect to GDH1 (such as Hawaii) have been affected more than 'normal' old lithosphere.

Thermal structure of the lithosphere

The difference between the predicted temperature structure of the lithosphere for GDH1 and PSM can be visualized using the asymptotic thermal structure for old lithosphere (Fig. 5), a linear geotherm, $T_m z/a = q_s z/k$. The age at which the geotherm becomes roughly linear corresponds to that at which the predicted heat flow becomes roughly uniform with age: ~100 Myr for GDH1 and ~150 Myr for PSM. The geotherm is easily interpreted graphically, because the physical properties are assumed to be uniform with depth. The surface heat flow constrains the slope of the geotherm. The geotherms for PSM and GDH1 can be compared directly because the models have the same conductivity. GDH1, with higher heat flow than PSM, has a steeper gradient and higher temperature at any depth. The observed subsidence constrains the area between the geotherm and $T = T_m$, which is proportional to $T_m a$ and thus to the heat lost as the plate cools from its initial temperature. The area between the geotherm and the depth axis is proportional to the heat remaining in the lithosphere, the vertical integral of temperature. For the linear gradient, the two areas are equal, so half of the initial heat has been lost. GDH1, derived for shallower bathymetry and hence less subsidence than PSM, has a smaller area between the geotherm and $T = T_m$. The model makes no predictions about the temperatures below the plate, often assumed to be a shallow adiabatic gradient, ~0.3 °C km⁻¹ (ref. 48). Given the simplifying assumptions and uncertainties, the temperatures at depth are in reasonable accord (±100 K) with those inferred from mineral physics⁴⁹.

Beyond the general properties that GDH1 is hotter at depth and has a thinner asymptotic thermal plate than PSM, we do not ascribe great significance to the details of the temperature structure, especially the basal temperature and the thermal structure in the lower portions of the plate. Plate models are phenomenological in that they are the simplest models (specified by the minimum number of parameters) that describe the general variation in depth and heat flow. These variations may reflect potentially significant effects not included in the model. For example, the basal temperature depends on whether the model incorporates radiogenic heat production, as illustrated by the

asymptotic geotherm $T(z) = q_a z/k - Hz^2/(2k)$ where H is the heat production per unit volume⁴⁸. Geotherm GDH1' (Fig. 5) shows that for a typically assumed heat production of 10^{-14} cal s cm⁻³ (4.184×10^{-8} W m⁻³)^{9,20}, the basal temperature would be reduced to 1,393 °C. Conversely, because in deriving PSM¹² it was assumed that an additional ~ 4 mW m⁻² of asymptotic heat flow was due to radioactivity, addition of this effect would give geotherm PSM', with a basal temperature of 1,409 °C. Because the basal temperature depends on the poorly known heat production, we have not complicated the GDH1 model by including this effect.

Other factors contribute to the uncertainty in the basal temperature. A depth dependent coefficient of thermal expansion²⁰ or conductivity⁵⁰ could have considerable effects. Because we use no heat-flow data in lithosphere younger than 50 Myr, we accept a boundary condition at the ridge which gives a heat-flow singularity that a more physical boundary condition can alleviate^{5,10,51}. Moreover, the temperatures at shallow depths in very young lithosphere are presumably lowered by hydrothermal circulation⁴⁴. The model includes only the variation in the depth due to the cooling lithosphere¹⁶, whereas large- and small-scale convection^{2,3,52}, temperature variations within the mantle, and phase changes⁵³ may also yield depth effects. The model does not incorporate time-dependent effects apart from cooling, as might be associated with processes like the proposed Cretaceous superplume⁵⁴.

Determination of the magnitude of effects not included in the plate model is challenging because the model fits the primary surface observables, depth and heat flow, reasonably well. A final point worth noting is that in the plate model, the ridge and basal temperatures are treated as the same for simplicity. These

temperatures are presumably different, in that the basal temperature represents the heat addition to old lithosphere. If this addition occurs through mantle plumes, the basal temperature would be expected to be several hundred degrees higher than the ridge temperature⁵⁵. This may be the reason why the basal temperature exceeds that ($\sim 1,300$ °C) inferred for the ridge temperature from the thickness of oceanic crust^{61,62}.

In addition to the surface observables, the thermal evolution of the lithosphere seems to give rise to variations with age of properties at depth, as inferred from surface-wave velocities^{20,56,57}, effective elastic thicknesses^{21,22} and the maximum depth of seismicity^{23,58}. It is hard to use these variations to discriminate between grossly similar thermal models like PSM and GDH1 (Fig. 6) because the predicted temperatures are similar at shallow depths, especially for young lithosphere. The inferred temperatures of ~ 400 °C associated with the effective elastic thickness²¹ (except for the South Pacific super-swallow⁵⁹) and of ~ 750 °C corresponding to the deepest intraplate seismicity²³ occur at similar depths for PSM and GDH1, except for old lithosphere. More detailed comparison of these data with predicted temperature structures requires rheological models and assumptions about the relevant strengths. The velocity structure inferred from surface-wave dispersion might be more diagnostic because the low-velocity zone occurs at greater depth, where predicted temperatures differ more. Laboratory experiments⁶⁰ indicate that the velocity should begin to decrease at a depth where the temperature is $\sim 95\%$ of the solidus temperature. Unfortunately, even with this assumption the velocity structure is difficult to relate in detail to thermal models because of the limited vertical and age resolution of the dispersion data^{20,56,57}. For the velocity structure shown, however, the inception of the low-velocity zone occurs at shallower depths in older lithosphere than would be expected for the present solidus data⁶⁰, either for PSM or the hotter GDH1.

The GDH1 model provides a convenient reference for the average variation in depth and heat flow with lithospheric age. It can be used in studying these surface observables, regardless of how well the plate model actually represents the thermal structure of the lithosphere. If a plate model adequately represents the geotherm, GDH1 implies that the lithosphere is hotter at depth and thinner than previously assumed. □

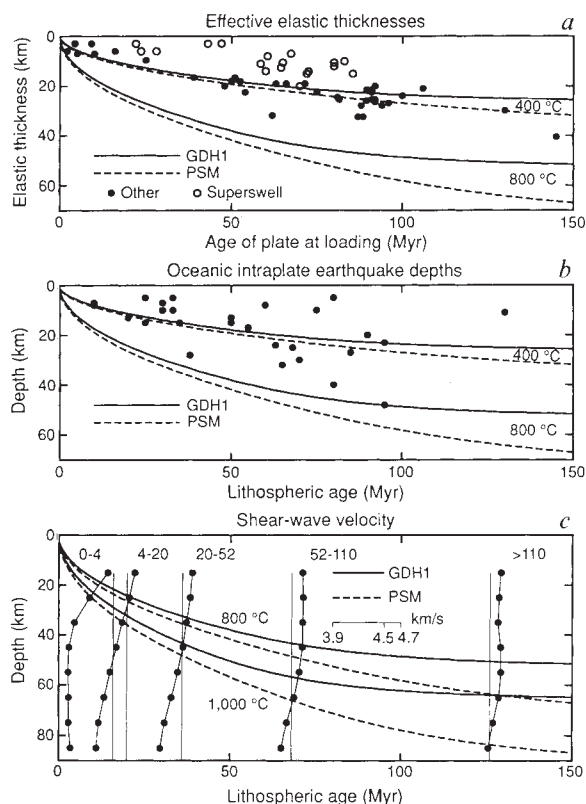


FIG. 6 Comparison of the isotherms as a function of age for the PSM and GDH1 models to three data sets^{22,23,56} whose variation with age is consistent with cooling of the lithosphere. Except for the oldest lithosphere, the isotherms corresponding to the effective elastic thickness (a) and deepest intraplate seismicity (b) are similar for the two models. The difference between the temperatures for the low-velocity zone (c) is greater. The vertical line for the velocity structure in each age range (for example 0–4 Myr) corresponds to 4.5 km s^{-1} .

Received 30 January; accepted 18 August 1992.

1. Parsons, B. & Richter, F. M. in *The Sea* 7 (ed. Emiliani, C.) 73–117 (Wiley, New York, 1981).
2. Jarvis, G. T. & Peltier, W. R. in *Mantle Convection* (ed. Peltier, W. R.) 479–594 (Gordon and Breach, New York, 1989).
3. O'Connell, R. J. & Hager, B. H. in *Physics of the Earth's Interior, Proc. int. Sch. Phys. "Enrico Fermi"* (eds Dziewonski, A. M. & Boschi, E.) 270–317 (North-Holland, Amsterdam, 1980).
4. Chapman, D. S. & Pollack, H. N. *Earth planet. Sci. Lett.* **28**, 23–32 (1975).
5. Slater, J. G., Jaupart, C. & Galson, D. *Rev. Geophys. Space Phys.* **18**, 269–311 (1980).
6. Davies, G. F. *Rev. Geophys. Space Phys.* **18**, 718–722 (1980).
7. Von Herzen, R. P. & Uyeda, S. *J. geophys. Res.* **68**, 4219–4250 (1963).
8. Langseth, M. G., Le Pichon, X. & Ewing, M. *J. geophys. Res.* **71**, 5321–5355 (1966).
9. Slater, J. G. & Francheteau, J. *Geophys. J. R. astr. Soc.* **20**, 509–542 (1970).
10. Davis, E. E. & Lister, C. R. B. *Earth planet. Sci. Lett.* **21**, 405–413 (1974).
11. McKenzie, D. P. *J. geophys. Res.* **72**, 6261–6273 (1967).
12. Parsons, B. & Slater, J. G. *J. geophys. Res.* **82**, 803–827 (1977).
13. Anderson, R. N. & Hobart, M. A. *J. geophys. Res.* **81**, 2968–2989 (1976).
14. Crough, S. T. *Phys. Earth planet. Inter.* **14**, 365–377 (1977).
15. Schubert, G., Froidevaux, C. & Yuen, D. A. *J. geophys. Res.* **81**, 3525–3540 (1976).
16. Parsons, B. & McKenzie, D. P. *J. geophys. Res.* **83**, 4485–4496 (1978).
17. Heestand, R. L. & Crough, S. T. *J. geophys. Res.* **86**, 6107–6114 (1981).
18. Schroeder, W. *J. geophys. Res.* **89**, 9873–9883 (1984).
19. Davies, G. F. *J. geophys. Res.* **93**, 10481–10488 (1988).
20. Forsyth, D. W. *Tectonophysics* **38**, 89–118 (1977).
21. Bodine, J. H., Steckler, M. S. & Watts, A. B. *J. geophys. Res.* **86**, 3695–3707 (1981).
22. Calmant, S., Francheteau, J. & Cazenave, A. *Geophys. J. int.* **100**, 59–67 (1990).
23. Wiens, D. A. & Stein, S. *J. geophys. Res.* **88**, 6455–6468 (1983).
24. Renkin, M. L. & Slater, J. G. *J. geophys. Res.* **93**, 2919–2935 (1988).
25. Slater, J. G. & Wixon, L. in *The Geology of North America: The Western North Atlantic Region* (eds Vogt, P. R. & Tucholke, B. E.) 257–270 (Geol. Soc. Am., Boulder, 1986).
26. Louden, K. E., Wallace, D. O. & Courtney, R. C. *Earth planet. Sci. Lett.* **83**, 109–122 (1987).
27. Lister, C. R. B., Slater, J. G., Davis, E. E., Villinger, H. & Nagihara, S. *Geophys. J. int.* **102**, 603–630 (1990).
28. Von Herzen, R., Cordery, P. M. J., Detrick, R. S. & Fang, C. *J. geophys. Res.* **94**, 13783–13799 (1989).
29. Stein, C. & Abbott, D. *J. geophys. Res.* **96**, 16083–16100 (1991).
30. Louden, K. E. in *Handbook of Seafloor Heat Flow* (eds Wright, J. A. & Louden, K. E.) 325–485 (CRC, Boca Raton, 1989).
31. Uyeda, S. in *Island Arcs, Deep-sea Trenches and Back-arc Basins, Maurice Ewing Ser. 1* (eds Talwani, M. & Pitman, W. C. III) 1–14 (Am. Geophys. Un., Washington DC, 1977).

32. Langseth, M. G., Lachenbruch, A. H. & Marshall, B. V. in *The Geology of North America: The Arctic Ocean Region* (eds Grantz, A., Johnson, L. & Sweeney, J. F.) 133–152 (Geol. Soc. Am., Boulder, 1990).
33. Kaula, W. M. & Phillips, R. J. *Geophys. Res. Lett.* **8**, 1187–1190 (1981).
34. Marty, J. C. & Cazenave, A. *Earth planet. Sci. Lett.* **94**, 301–315 (1989).
35. Colin, P. & Fleitout, L. *Geophys. Res. Lett.* **17**, 1961–1964 (1990).
36. Smith, W. H. F. thesis, Columbia Univ. (1990).
37. Detrick, R. S. & Crough, S. T. *J. geophys. Res.* **83**, 1236–1244 (1978).
38. Von Herzen, R. P., Detrick, R. S., Crough, S. T., Epp, D. & Fehn, U. *J. geophys. Res.* **87**, 6711–6723 (1982).
39. Courtney, R. C. & White, R. S. *Geophys. J. R. astr. Soc.* **87**, 815–867 (1986).
40. Robinson, E. M. & Parsons, B. J. *geophys. Res.* **93**, 3144–3156 (1988).
41. Liu, M. & Chase, C. G. *J. geophys. Res.* **94**, 5571–5584 (1989).
42. Woods, M. T., Leveque, J.-J. & Okal, E. A. *Geophys. Res. Lett.* **18**, 105–108 (1991).
43. Sleep, N. H. & Wolery, T. J. *J. geophys. Res.* **83**, 5913–5922 (1978).
44. Lin, J. & Parmentier, E. M. *Geophys. J.* **96**, 1–22 (1989).
45. Cochran, J. R. *Geophys. J. R. astr. Soc.* **87**, 421–454 (1986).
46. Hayes, D. E. *J. geophys. Res.* **93**, 2937–2954 (1988).
47. Sleep, N. H. *J. geophys. Res.* **95**, 21983–21990 (1990).
48. Turcotte, D. L. & Schubert, G. *Geodynamics: Applications of Continuum Physics to Geological Problems* (Wiley, New York, 1982).
49. Akaogi, M., Ito, E. & Navrotsky, A. *J. geophys. Res.* **94**, 15671–15685 (1989).
50. Horai, K. & Susaki, J. *Phys. Earth planet. int.* **55**, 292–305 (1989).
51. Sleep, N. H. *J. geophys. Res.* **80**, 4037–4042 (1975).
52. Buck, W. R. & Parmentier, E. M. *J. geophys. Res.* **91**, 1961–1974 (1986).
53. Wood, B. J. & Yuen, D. A. *Earth planet. Sci. Lett.* **66**, 303–314 (1983).
54. Larson, R. L. & Olson, P. *Earth planet. Sci. Lett.* **107**, 437–447 (1991).
55. Sleep, N. H. A. *Rev. Earth planet. Sci.* **20**, 19–43 (1992).
56. Nishimura, C. & Forsyth, D. *Geophys. J.* **96**, 203–226 (1989).
57. Yu, G.-K. & Mitchell, B. *Geophys. J.* **57**, 311–342 (1979).
58. Chen, W.-P. & Molnar, P. *J. geophys. Res.* **88**, 4183–4214 (1983).
59. McNutt, M. K. & Fisher, K. M. in *Seamounts, Islands & Atolls*, *Geophysical Monograph #43* (eds Keating, B. H., Fryer, P., Batiza, R. & Boehler, G. W.) 25–34 (Am. Geophys. Un., Washington DC, 1987).
60. Sato, H., Sacks, I. S. & Murase, T. *J. geophys. Res.* **94**, 5689–5704 (1989).
61. Sleep, N. H. & Windsley, D. F. *J. Geol.* **90**, 363–379 (1982).
62. McKenzie, D. & Bickle, M. J. *J. Petrol.* **29**, 625–679 (1988).

ACKNOWLEDGEMENTS. We thank D. Abbott and Z. Wang for assistance in analysis of the heat-flow data, including determining site ages, W. Smith for a preprint, C. Bina, R. Carlson, R. Gordon, M. Langseth, R. O'Connell, T. Shoberg, N. Sleep and M. Woods for discussions, and R. Von Herzen for comments. The research was supported by the NSF, NASA and the Petroleum Research Fund of the American Chemical Society.

LETTERS TO NATURE

Direct measurement of the optical depth in a spiral galaxy

Raymond E. White III & William C. Keel

Department of Physics and Astronomy, University of Alabama,
Box 870324, Tuscaloosa, Alabama 35487-0324, USA

FROM a statistical analysis of nearly 9,400 spiral galaxies¹, Valentijn² has claimed that the disks of spirals are largely opaque. His argument derives from a lack of inclination dependence in the average surface brightness of the spirals, which if they were transparent would be brighter when seen edge-on than face-on. This statistically derived result is however vulnerable to several selection effects³, and seems to contradict the fact that the survey was successfully performed (as we live in a transparent spiral galaxy) as well as anecdotal examples of galaxies visible through other galaxies. We have tried to measure directly optical extinction in a number of spiral disks, using pairs in which a foreground spiral is backlit by another galaxy. In our best example substantial extinction of starlight occurs, but the extinction is mostly associated with the spiral features in which much of the light arises in the first place. This correlation of extinction with emission may account for both the transparency and apparent opacity of spiral galaxies.

The opacity of spiral disks bears on a variety of issues, including the nature of dark matter (if spirals are intrinsically more luminous than usually assumed, they would require less dark matter to contribute to the dynamical mass); galaxy star-formation rates (if more dust is available to absorb and reradiate starlight, starburst systems may be less extreme than originally thought); and the quasars redshift cutoff (due, perhaps, to the accumulation of intervening galaxies).

In Valentijn's study², the opacity of spiral disks was indicated by the lack of inclination dependence in their average surface brightness; if spiral disks were instead transparent, they would have higher surface brightness when seen edge-on than when viewed face-on. Burstein *et al.*³ recently discussed the observational selection effects inherent in such statistical studies; they attributed Valentijn's result to a selection effect associated with the diameter-limited nature of his sample. But their own statistical analysis, which was designed to avoid such selection effects, also suggested that spirals are optically thick.

We have adopted a direct, rather than statistical, approach to determine whether spiral disks are opaque. We have imaged several overlapping pairs of galaxies so that we may reconstruct, using symmetry, the amount of background light absorbed by the foreground spiral galaxies in the pairs. The ideal pair for this measurement consists of a symmetric face-on spiral in the foreground half-projected against half of a similar-sized, symmetric background elliptical (see Fig. 1). Both galaxies need to be highly symmetric so that their non-overlapping portions can

be used to estimate the superposed portions accurately, through either reflection or rotational symmetry.

Figure 1 illustrates the procedure for extracting optical depths from an idealized foreground 'spiral' disk projected against a background 'elliptical'. The surface brightness of the unsuperposed part of the spiral is S and that of the unsuperposed part of the elliptical is E , so the surface brightness in the superposed region is $S + Ee^{-\tau}$, where τ is the optical depth of the spiral. We estimate $e^{-\tau}$ by first subtracting the unprojected portion of the spiral (S) from its counterpart in the overlap region, then dividing the residuals by the unsuperposed portion of the elliptical (E). This results in a map of $e^{-\tau}$ over the projected parts of the foreground spiral. This technique is purely differential, so we do not require absolute intensity calibrations, only a linear detector.

Our cleanest case thus far is AM1316–241 (ref. 4), a foreground Sbc spiral projected against an elliptical galaxy (see the B (blue) image in Fig. 2a). The spiral and elliptical have galactocentric velocities of $10,365 \pm 30$ and $9,703 \pm 25$ km s⁻¹, respectively (D. F. de Mello Rabaça, personal communication). The velocity difference of 662 ± 39 km s⁻¹ is rather large for this to be a closely interacting bound pair. The catalogued surface brightness of the foreground spiral ($\mu_0^B = 21.36$ or $\mu_e^B = 23.07$) is 'typical', being within 0.2 magnitudes of the mean for all late-type galaxies with similar axial ratio (see Fig. 1 of ref. 2). The 4:1 ratio of the major and minor axes indicates the spiral is inclined $\sim 75^\circ$ from the plane of the sky, so there is a long optical path through the disk. Both galaxies have good symmetry, so the opacity can be accurately decomposed. We obtained B and I (infrared) images using the 2.1-m telescope of Kitt Peak National Observatory with a Tektronix 1,024 × 1,024-pixel CCD; exposure times were 30 minutes each.

The B/I colour ratio is shown in Fig. 2b, where it is obvious that the foreground galaxy is not opaque. Its disk is redder in the region backlit by the elliptical than in the corresponding unlit region, requiring that some (redder) background light shine through the disk (suffering additional reddening there). If the foreground spiral were opaque, there would be no colour gradient correlated with the surface brightness distribution of

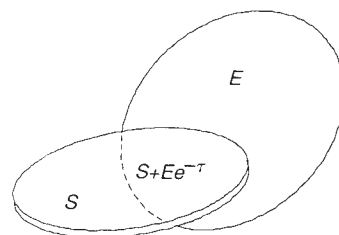


FIG. 1 Schematic of the ideal overlapping galaxy pair for the direct determination of the opacity of the foreground spiral.

TABLE 1 Apparent and scattering-corrected optical depths

	Arm	Interarm	Disk average
τ_B	1.4	0.3	0.7
τ_B^{sc}	1.45–1.7	0.3–0.4	0.75–1.0
τ_I	0.6	0.2	0.5
τ_I^{sc}	0.6	0.2	0.5–0.6

Scattering-corrected optical depths are for line-of-sight separations of $720 h_{50}^{-1}$ kpc and $72 h_{50}^{-1}$ kpc, respectively.

the background elliptical. The differential reddening is greatest in the projected spiral arm and least in the interarm regions.

In Figs 2c and d are B and I images of 'apparent' $e^{-\tau}$ constructed in the manner described above. Corrections may be needed for the scattering of off-axis background elliptical light into the line of sight, which would cause the inferred optical depths to be underestimated. In the B band, the elliptical light is reduced by a median factor of $e^{-\tau_B} \approx 0.25$ in the arm, so the optical depth is $\tau_B \approx 1.4$. In the interarm region the spiral is nearly transparent, with a median value of $\tau_B \approx 0.3$. The optical depth of the projected arm is less in the I band, where the median is $\tau_I \approx 0.6$. In the interarm region, the disk is nearly transparent in I as well, with a median $\tau_I \approx 0.2$. Table 1 summarizes these optical depths. Face-on optical depths would be ~ 4 times smaller if the absorbing material is smoothly distributed in the disk.

The optical depth is reasonably well determined from the outer arm, $\sim 28''$ from the centre of the spiral, to $\sim 14''$ from the centre, where the reflection symmetry breaks down. We calculate the Holmberg radius (where the blue surface brightness diminishes to $\mu_B = 26.5$ mag arcsec $^{-2}$) to be $44''$, in good agreement with the value of $47''$ obtained by Lauberts and Valentijn¹. The prominent dust lane is then 0.64 Holmberg radii out, and we measure the optical depth in to 0.3 Holmberg radii. The area in which we have good optical depths encompasses $\sim 1,500$ pixels (our scale is $0.31''$ per pixel). The median fraction of elliptical light filtering through the (measurable) disk is ~ 0.5 in

the blue and ~ 0.6 in the infrared, so the median optical depths are $\tau_B \approx 0.7$ and $\tau_I \approx 0.5$.

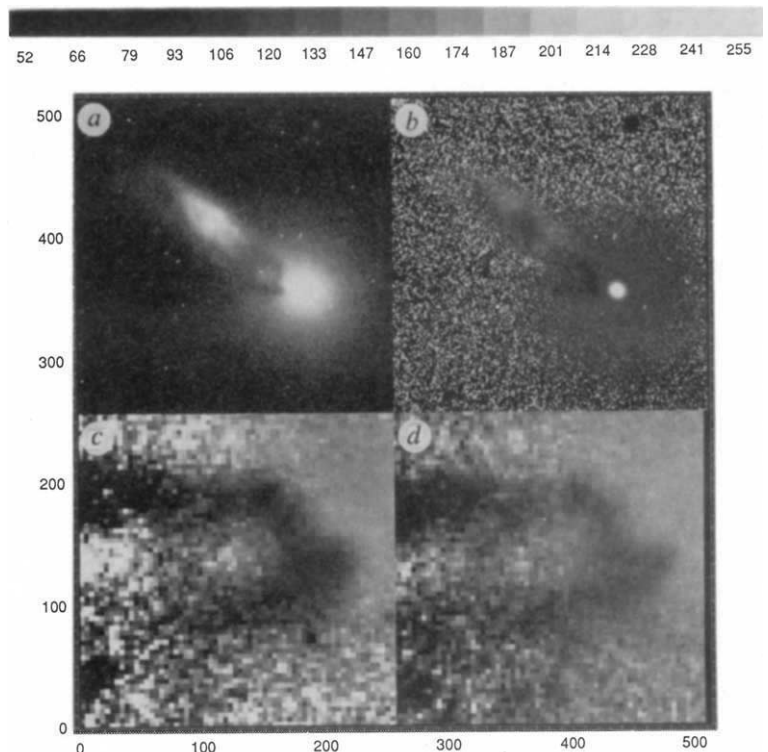
Apart from possible scattering effects, the observational errors in our extinction estimates are dominated by systematics (due to asymmetries in the foreground spiral) rather than photon statistics. The morphological similarity of the absorbing arm in the $e^{-\tau}$ plots (Fig. 2c and d) to that in the colour ratio image (Fig. 2b) shows that these systematic errors do not dominate our results in the region demarcated above. Nonetheless, to quantify the effect of systematic errors upon our opacity estimates, let us assume that the actual surface brightness of the foreground spiral in the overlap region is S , and that of the background elliptical is E . We denote our estimates of S and E , derived from the non-overlap regions, as S' and E' . Our estimate of $e^{-\tau}$, denoted by $e^{-\tau'}$, is then related to the actual value through

$$e^{-\tau'} = \frac{E}{E'} e^{-\tau} + \left[\frac{S}{S'} - 1 \right] \frac{S'}{E'}$$

E/E' is within a few per cent of unity in this case, so the bulk of the systematic error comes from asymmetry in the foreground spiral. But the spiral asymmetry factor $(S/S' - 1)$ is weighted by the surface brightness ratio S'/E' . In the regions of interest, $S'/E' \approx 0.21$ in B and 0.16 in I, so even a sizeable uncertainty in S will have only a small effect on our estimate of $e^{-\tau}$. To estimate S/S' , we divided the half-spiral in the overlap region by a reflection of its non-overlapping half. Most of the resulting pixel values in an arm region close to the reflection axis (away from contamination by the background elliptical) are near unity: in the B image, 2/3 of the pixels have values between 0.7 and 1.5, and in I, 2/3 are between 0.8 and 1.2. Thus we estimate our systematic errors in $e^{-\tau}$ to be ≤ 0.1 in B and ~ 0.03 in I.

If the absorbing material consists of optically thin material, off-axis light from the extended background source may scatter into the line of sight, causing the optical depth to be underestimated. (To the extent that its distribution shows reflection symmetry, internal scattering of the spiral's own light does not affect our results because it has been subtracted off). If the absorbing material is composed of optically thick clumps of low albedo, scattering would have negligible effect on our optical

FIG. 2 AM1316–241: a, B image (north is to the right, east to the top). b, B/I colour ratio. c, Distribution of $e^{-\tau_B}$, where τ_B is the blue optical depth (image rotated 39° counterclockwise relative to a and b). d, same as c, but in the I band.



depth estimates. We assess the maximum possible effects of scattering by assuming that the absorbing material is optically thin and smoothly distributed. We take dust grain properties (albedos and scattering asymmetry parameters) in B and I from Bruzual *et al.*⁵ and follow their prescription for calculating the effects of scattering, using the Henyey and Greenstein⁶ scattering phase function. We consider points in the outer arm and interarm regions and calculate optical depth corrections for a variety of assumed line-of-sight separations between the spiral and elliptical. The absence of tidal distortion in the galaxies' isophotes provides the minimum plausible separation, because the galaxies are then not physically overlapping. The sum of their measurable radii corresponds to a minimum line-of-sight separation of ~ 240 pixels ($72 h_{50}^{-1}$ kpc, where h_{50} is the Hubble constant in units of $50 \text{ km s}^{-1} \text{ Mpc}^{-1}$). We also consider separations up to 10 times larger. The scattering-corrected median optical depths are shown in Table 1. Evidently, scattering effects are negligible in I, but may be as large as 30–40% in B if the galaxies are as close as ~ 70 kpc (and the dust is not in optically thick clumps). The effects diminish rapidly with increasing separation, however: if the separation exceeds only $\sim 150 h_{50}^{-1}$ kpc, the B corrections are $\leq 15\%$. In any case, our main result regarding the segregation of absorbing material remains unaltered.

Because the observed extinction is largely confined to the spiral arms, it is correlated with the distribution of visible light, which is also concentrated in the spiral arms. In the region where we have measured the B optical depths, 50% of the light arises from only 25% of the area. In I, 50% of the light comes from 30% of the area. Spirals that are more nearly face-on tend to show an even greater concentration of light in the arms. This

correlation of internal extinction with the emission from spiral disks can reconcile two seemingly contradictory observations: galaxies are seen through other galaxies (implying that spirals are optically thin); yet the average surface brightness of spirals is independent of inclination (implying that spirals are optically thick). Although 'patchy' absorption may allow spirals to be partially transparent, random patchiness will not lead to the general optical thickness of spirals noted by Valentijn² and Burstein *et al.*³ (unless the covering fraction of patches is so large that few galaxies will be partially transparent). But if disk extinction is concentrated in the brightest regions, as our observations suggest, spirals can appear to be both optically thick (with surface brightness independent of inclination) and partially transparent (in interarm regions, at least). The Sun's location in the outskirts, rather than the middle, of the Galaxy's Orion arm is fortunate for observers of extragalactic phenomena. $\square$

Received 22 November 1991; accepted 24 July 1992.

1. Lauberts, A. & Valentijn, E. A. *The Surface Photometry Catalogue of the ESO-Uppsala Galaxies* (European Southern Observatory, Munich, 1989).
2. Valentijn, E. *Nature* **346**, 153–155 (1990).
3. Burstein, D., Haynes, M. P. & Faber, S. M. *Nature* **353**, 515–521 (1991).
4. Arp, H. C. & Madore, B. F. *A Catalogue of Southern Peculiar Galaxies and Associations* (Cambridge Univ. Press, 1987).
5. Bruzual, G., Gladis Magris, C. & Calvet, N. *Astrophys. J.* **333**, 673–688 (1988).
6. Henyey, L. G. & Greenstein, J. L. *Astrophys. J.* **93**, 70 (1941).

ACKNOWLEDGEMENTS. Kitt Peak National Observatory, NOAO, is operated by AURA, Inc., under contract with the NSF. We thank D. F. de Mello Rabaça for obtaining a spectrum of AM1316-241 at the European Southern Observatory. This work was supported by the State of Alabama and the NSF (EPSCoR).

The case against a massive black hole at the Galactic Centre

R. H. Sanders

Kapteyn Astronomical Institute, University of Groningen,
9700 AV Groningen, The Netherlands

MANY of the compact near-infrared sources observed in the central parsec of our Galaxy have been shown to be very massive, luminous blue stars with strong helium lines and fast winds, perhaps similar to Wolf-Rayet stars^{1–3}; and the complex near-infrared source IRS16, which appears to be the central concentration of such sources, has now been shown³ to consist of about 20 luminous blue stars, not 100 main-sequence stars, as had been suggested earlier⁴. These young blue stars, rather than any object associated with the nearby compact radio source SgrA* (ref. 8), may be the predominant source of ionizing radiation¹ and hydrodynamic activity^{5–7} in the inner two parsecs of the Galaxy. An earlier argument⁴ that the distribution of IRS16 components precludes the existence of a million-solar-mass black hole at or near SgrA* remains valid even if there are fewer than two dozen stars in IRS16. Here I argue further that the presence of young stars in this region is incompatible with the presence of a giant black hole or any other centrally condensed mass, because tidal forces would have inhibited star formation.

It has been argued that the patchy and asymmetric distribution of IRS16 components with respect to SgrA* is inconsistent with the short orbit times in the near vicinity of a massive black hole⁴. More recent observations reveal that IRS16 consist of about 20 very luminous blue stars³ instead of 100 main-sequence B stars as supposed in ref. 4, and that the IRS16 stars are apparently the brightest members of a larger cluster of young mass-losing stars extending to a distance of 0.5 pc from SgrA* (refs 1, 2). These facts do not eliminate the problem that the apparent centroid of the IRS16 group lies distinctly to the east of SgrA*

by about one second of arc^{3,4}. The orbit time about a massive central object would be

$$t_d \approx 1,000 \left(\frac{M_h}{10^6 M_\odot} \right)^{-0.5} \left(\frac{r}{0.1 \text{ pc}} \right)^{1.5} \text{ years}$$

where M_h is the mass of the hole, $M_\odot$ is the mass of the Sun and r is radius of orbit. The timescale for orbital phase mixing is comparable to t_d . The typical projected distance of an IRS16 star from SgrA* is about 0.1 pc. If the true distance is comparable, then even though the stars are only a few million years old, the IRS16 group should be randomly distributed about the black hole. The probability that the 20 most luminous stars comprising the central concentration of the 0.5 pc cluster of young stars all lie to the east of SgrA* at the present epoch (as observed) is less than 10^{-4} .

It seems unlikely that the true distance of this unique grouping of young stars from SgrA* is much larger than the projected distance, although this would avoid the problem of orbital mixing. The actual distance would have to be larger than about 2 pc; otherwise, the IRS16 group would disperse over the inner 40" in less than 10^6 years. Such a large distance would be inconsistent with the observational evidence that these stars are the source of the wind impinging on the neutral gas annulus at 2 pc (refs 1, 8).

The luminous IRS16 stars, as well as the other blue stars in the central 0.5 pc, were probably formed near where they are observed now. If they were formed significantly further out ($r > 1$ pc) on radial orbits which carry them into the centre, then, again because of the short orbit times, it would be improbable to find them all clustered in the inner few tenths of a parsec. Two-body relaxation or dynamical friction⁹, which might drag massive stars inward, would operate on a much longer timescale ($> 10^8$ years) than the age of the stars.

Star formation in the near tidal field of the hypothetical black hole can proceed only under extreme conditions. In particular, the particle number density of the protostellar clouds must exceed

the Roche limit of

$$\rho > \frac{M_h}{r^3} \approx 4 \times 10^{10} \frac{M_h}{10^6 M_\odot} \left(\frac{0.1 \text{ pc}}{r} \right)^3 \text{ cm}^{-3}$$

Inferred electron densities in the ionized gas streamers observed in the central parsec do not exceed 10^5 cm^{-3} . Phinney¹⁰ has argued that the post-shock density in colliding gas streams might be as high as 10^9 cm^{-3} . Even so, the possibility of star formation within 0.5 pc of a $10^6 M_\odot$ black hole is marginal at best, unless gas densities in the central parsec were very much higher a few million years ago.

A mass model containing no central mass concentration does not conflict with the present observations. The mass is entirely in the old stellar population with a density distribution given by

$$\rho = 1.9 \times 10^6 [1 + (3r_{\text{pc}})^2]^{-0.9} M_\odot \text{ pc}^{-3} \quad (1)$$

where r_{pc} is the radial distance to the centre in parsecs. This is consistent with the mean radial distribution of the $2.2 \mu\text{m}$ intensity between 1 and 30 pc (ref. 11). In the context of this model the continued increase in the $2.2\text{-}\mu\text{m}$ surface brightness into one or two seconds of arc would be due to contamination of the emission by the central cluster of young luminous stars^{1,12}. This effect could also cause the observed decrease in the strength of the CO band head in the inner 10 arcsec (ref. 13).

At radii much larger than the $\frac{1}{3}$ pc core radius, the total interior mass, M_r , assuming spherical symmetry, is

$$M_r = 2.7 \times 10^6 (r_{\text{pc}})^{1.2} M_\odot \quad (2)$$

which is consistent with several kinematic tracers of the galactic bulge mass distribution^{8,14} and with a reasonable value for the bulge mass-to-light ratio¹⁵. In the very central region the volume integral of equation (1) yields a total stellar mass of $5.2 \times 10^6 M_\odot$ inside a radius of 2 pc and $2.1 \times 10^6 M_\odot$ inside 1.1 pc. These values agree with those implied by the rotating neutral gas disc observed between 2 pc and 8 pc (ref. 8), and by the rotating ring of ionized gas observed in hydrogen radio recombination lines between 1.1 and 1.6 pc (ref. 16, with results scaled to a Galactic Centre distance of 8.5 kpc). It is this second observation which sets the value of the core radius.

The model is also consistent with the line-of-sight velocity dispersion of stars determined from the measured radial velocities of individual late-type giants ($70\text{--}80 \text{ km s}^{-1}$) between 0.3 and 4 pc (ref. 17). But the stellar velocity dispersion determined from broadening of the CO band head at $2.3 \mu\text{m}$ apparently increases to about 125 km s^{-1} in the inner half parsec^{13,18}. This can arise from a mildly anisotropic velocity distribution, weighted toward more radial orbits, and does not necessarily require a central mass concentration (R.H.S., manuscript in preparation). Indeed, if the increase in velocity dispersion were due to a massive object, then the required mass, $\sim 4 \times 10^6 M_\odot$, would exceed that which is inside the rotating ring of ionized gas at 1.1 pc (ref. 16).

In principle, the measured velocity dispersion³ of 17 blue luminous stars inside 0.5 pc ($\sim 130 \text{ km s}^{-1}$) should also be a kinematic tracer of the mass distribution in the central core of the Galaxy. In fact, the mean systematic redshift of the stars ($\sim 80 \text{ km s}^{-1}$), apparently due to aspects of the helium line

formation in the expanding atmospheres of these stars^{1,2}, implies that the measured value can only be considered as an upper limit. Even this upper limit does not support the suggestion that the true dispersion continues to rise within a few tenths of a parsec of SgrA*.

In the context of this model (equation (1)), the luminous blue stars giving rise to IRS16 would have formed inside a core with a roughly constant density of old stars. The gravitational potential here is essentially that of a simple harmonic oscillator, and there is no tidal limit to the density of protostellar clouds; on the contrary, the tidal field actually contributes to gravitational collapse. For the mass distribution implied by equation (1), the tidal field inhibits gravitational collapse only at radii larger than 0.6 pc, corresponding to the extent of the central cluster of young stars.

In the absence of a massive black hole, the centroid of the IRS16 group would correspond to the centre of this potential well. The evidence for hydrodynamic activity in the central parsec, such as the cometary tail of the late-type giant^{7,19} and the mini-cavity⁷, must all be attributed to the observed winds from these newly formed stars. It might be that an inactive massive black hole in a normal galactic nucleus, such as that of M31 (ref. 20), actually inhibits these mild forms of 'activity' by the tidal suppression of star formation in the central few parsecs.

Although this model can explain the observations without evoking a massive black hole, it does leave the unique compact radio source near the centre unexplained. One of the principal arguments in favour of SgrA* being a very massive object is its immobility. Proper-motion studies²¹ show that this source has a transverse velocity, after subtracting solar motion, of less than 40 km s^{-1} . Presumably, therefore, SgrA* is a massive body sitting in the bottom of the potential well of the Galaxy.

This argument is not compelling. Suppose that SgrA* is a black hole having a moderate mass of $10 M_\odot$, the sort of object which might be formed in the normal course of stellar evolution. If the black hole has been around for more than 10^9 years it would have come into thermal equilibrium with the lower-mass core stars, which implies that its kinetic energy should be comparable. Because the core stars in the harmonic potential well have a r.m.s. space velocity of about 100 km s^{-1} , the black hole would have a typical space velocity of 30 km s^{-1} , below the upper limit on the transverse velocity. Its maximum radial excursion from the centre of the potential well would be $\sim r_{\text{core}}/\sqrt{10} \approx 0.1 \text{ pc}$, which is consistent with the observed projected separation of SgrA* from the centroid of the IRS16 cluster ($\sim 0.05 \text{ pc}$). This can be rearranged to set an upper limit on the mass of an object associated with SgrA*. If the true displacement of SgrA* from the centre of the potential well is equal to the observed displacement from the IRS16 centroid, then its mass cannot exceed $50 M_\odot$.

It is difficult to understand the presence and distribution of young stars at the Galactic Centre if SgrA* is identified with a massive black hole. But these nonexistence arguments are indirect and can never be absolutely conclusive. The essential test is dynamical; radial velocity or proper-motion measurements of the IRS16 components and SgrA* remain of the utmost importance. $\square$

Received 22 April; accepted 16 July 1992.

- Allen, D. A., Hyland, A. R. & Hillier, D. J. *Mon. Not. R. astr. Soc.* **244**, 706–713 (1990).
- Krabbe, A., Genzel, R., Drapatz, S. & Rotaciuc, V. *Astrophys. J.* **382**, L19–L22 (1991).
- Eckart, A. et al. *Nature* **355**, 526–529 (1992).
- Allen, D. A. & Sanders, R. H. *Nature* **319**, 191–194 (1986).
- Gately, I. et al. *Mon. Not. R. astr. Soc.* **222**, 299–306 (1986).
- Yusef-Zadeh, F., Morris, M. & Ekers, R. D. *IAU Symp. 136, The Center of the Galaxy* (ed. Morris, M.) 443–451 (Kluwer, Dordrecht, 1989).
- Yusef-Zadeh, F. & Morris, M. *Astrophys. J.* **371**, L59–L62 (1991).
- Genzel, R. & Towns, C. H. A. *Rev. Astr. Astrophys.* **25**, 377–423 (1987).
- Chandrasekhar, S. *Principles of Stellar Dynamics* (Dover, New York, 1960).
- Phinney, E. S. *IAU Symp. 136, The Center of the Galaxy* (ed. Morris, M.) 543–553 (Kluwer, Dordrecht, 1989).
- Becklin, E. E. & Neugebauer, G. *Astrophys. J.* **151**, 145–161 (1968).

- Reike, G. H. & Lebofsky, M. J. *The Galactic Center*, AIP Conf. Proc. **155** (ed. Backer, D. C.), 91–94 (American Institute of Physics, New York, 1987).
- Sellgren, K., McGinn, M. T., Becklin, E. E. & Hall, D. N. B. *Astrophys. J.* **359**, 112–120 (1990).
- Lindqvist, M., Habing, H. J. & Winnberg, A. *Astr. Astrophys.* **259**, 118–127 (1992).
- Kent, S. M. *Astrophys. J.* **387**, 181–188 (1992).
- Schwarz, U. J., Bregman, J. D. & van Gorkom Astr. *Astrophys.* **215**, 33–47 (1989).
- Reike, G. H. & Reike, M. J. *Astrophys. J.* **330**, L33–L37 (1988).
- McGinn, M. T., Sellgren, K., Becklin, E. E. & Hall, D. N. B. *Astrophys. J.* **338**, 824–840 (1989).
- Yusef-Zadeh, F. & Melia, F. *Astrophys. J.* **385**, L41–L44 (1992).
- Kormendy, J. *Astrophys. J.* **325**, 128–141 (1988).
- Backer, D. C. & Sramek, R. A. *The Galactic Center* (ed. Backer, D. C.) 163–165 (American Institute of Physics, New York, 1986).

ACKNOWLEDGEMENTS. I thank D. Allen for discussions of these ideas at an earlier stage and U. Schwarz and B. McNamara for useful comments on the initial manuscript.

Friction measurements on phase-separated thin films with a modified atomic force microscope

R. M. Overney*, E. Meyer*, J. Frommer*[‡],
D. Brodbeck*[‡], R. Lüthi*, L. Howald*,
H.-J. Güntherodt*, M. Fujihira†,
H. Takano† & Y. Gotoh†

* Physics Institute, University of Basel, Klingelbergstrasse 82,
CH-4056 Basel, Switzerland

† Department of Biomolecular Engineering, Tokyo Institute of Technology,
Nagatsuta, Midori-ku, Yokohama 227, Japan

THE study of chemical phase separation in multicomponent thin organic films typically involves the addition of a dye which is selectively more soluble in one of the phases, thereby making it possible to probe the domain structures by fluorescence microscopy^{1–4}. The resolution of this approach is generally limited to tens of micrometres. The atomic force microscope, on the other hand, has recently proved useful for imaging organic thin films down to the atomic scale^{5–9}, but this technique provides details of the overall film topography, rather than the chemical composition. Here we show that the recently developed friction force microscope^{10–13}, which simultaneously measures both the normal and lateral forces on the scanning tip, can be used to image and identify compositional domains with a resolution of ~ 5 Å. Although the topography of the individual domains can be imaged with a standard atomic force microscope, it is the additional information provided by the friction measurement that allows them to be chemically differentiated.

Scanning probe microscopes such as the scanning tunnelling microscope¹⁴ (STM) and atomic force microscope⁵ (AFM) have been used to analyse surfaces on many levels, from molecular arrangements at interfaces to electronic structures in semiconductors⁹. These methods are rather insensitive to chemical composition; with the STM, electronic states around the Fermi energy are probed, whereas AFM images are interpreted as maps of the total charge density^{15,16}. Some elemental characterization has been possible with scanning tunnelling spectroscopy on samples such as GaAs¹⁷, adsorbates on semiconductors¹⁸ and metals¹⁹. For more detailed information, core levels have to be sensed, and this is not possible with the STM^{18,20}.

Frictional forces can influence AFM image contrast when it is operating in the repulsive contact mode^{10,11}. Those measurements show that frictional forces vary on an atomic scale and that a friction force coefficient for a tungsten tip on a graphite sample can be determined¹⁰. These observations led to the construction of a friction force microscope (FFM) where both the normal and lateral forces are measured simultaneously but separately over homogeneous surfaces^{11–13}. In this instrument, the bending and torsion of a cantilever-type spring are correlated with the normal and lateral forces exerted by the sample, while the fine resolution of an AFM is maintained. Here, we extend the use of the FFM to samples composed of distinctly different chemical species and demonstrate that the FFM can distinguish between these components.

From macroscopic measurements it is well known that friction depends strongly on material. Various types of apparatus have been used to measure frictional properties as a function of temperature, sliding velocity and normal load^{21–23}. The FFM represents a simplified device for experiments on sliding friction under boundary conditions: an idealized single asperity on a thin multilayer film. With the FFM the lateral (friction) forces acting on the probing tip can be measured over a localized

contact area of a few square nanometres between probing tip and sample.

To demonstrate that on the nanometre scale, frictional properties are sensitive to changes in surface properties on chemical modification^{24,25}, we have chosen Langmuir–Blodgett (LB) films because of their solid-like behaviour, their use in previous work on boundary lubrication²¹ and their well-defined, non-random structure²⁶, so that the results would be straightforward to interpret. But the FFM technique is applicable to a wide range of materials, for the study of material-specific friction and as a means of identifying materials on this very small scale from their different frictional responses. Few, if any, other techniques can probe material properties directly and nondestructively on so small a scale with such good resolution (~ 5 Å). Its usefulness has been demonstrated on layered materials such as mica¹² and graphite¹⁰. Here we demonstrate that the technique can be used to distinguish between components in a multicomponent system, as we assess not only the differential friction between two molecular classes (fluorocarbon and hydrocarbon), but also that between organic adsorbate and inorganic (silicon) substrate.

Figure 1a shows a $5.0 \times 5.0 \mu\text{m}^2$ top-view AFM image of a 50:50 molar mixture of arachidic acid and partially fluorinated carboxylic acid bound ionically to a common cationic polymer (see Fig. 2). Round island-like structures are observed, 100 nm to 1 μm in diameter and ~ 1.6 nm in height above the surrounding film. We assign the higher circular domains in Fig. 1a to the hydrocarbons and the surrounding flat film to the fluorocarbons, partly on the basis of their different molecular lengths and partly because of their different frictional behaviours (see below). With molecular lengths of ~ 2.5 nm and ~ 2.0 nm for the arachidic acid and the fluorocarbon-terminated acid,

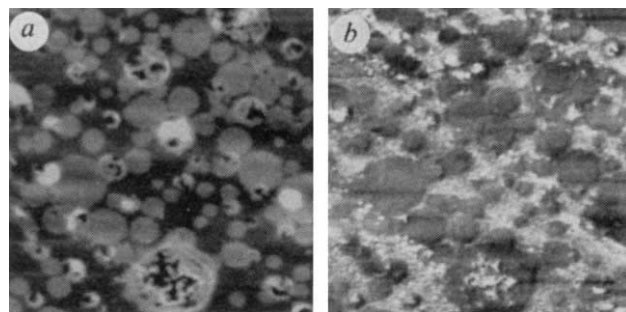


FIG. 1 The LB films are prepared as described elsewhere^{37,38}. Briefly, a 1:1 molar mixture of perhydro arachidic acid ($\text{C}_{19}\text{H}_{39}\text{COOH}$) and partially fluorinated carboxylic acid, ($\text{C}_9\text{F}_{19}\text{C}_2\text{H}_4\text{OC}_2\text{H}_4\text{COOH}$), is introduced in organic solvent to an aqueous subphase containing poly(4-vinyl-*N*-methylpyridinium) cationic polymer. The complex of anionic surfactants and cationic polymer is transferred by the LB technique onto a hydrophobized silicon (100) surface to form a bilayer system. Pretreatment of the silicon surface is explained elsewhere³⁹. The deposition of the LB films is carried out in a Langmuir trough at room temperature. The FFM cantilevers used here are micro-fabricated from Si_3N_4 and have normal and torsion spring constants of 0.2 N m^{-1} and 200 N m^{-1} , respectively. The resolution of the instrument is 0.1 Å in the normal direction and 10^{-10} N in the friction force. The measurements are done at minimal normal load, of the order of 10^{-9} – 10^{-8} N. **a**, AFM image ($5 \times 5 \mu\text{m}^2$) of the surface of a bilayer prepared from a mixture of the fluorocarbon and hydrocarbon carboxylates (1:1 molar). The image represents the topography. The circular domains are assigned to the hydrocarbon component and the surrounding flat film to the partially fluorinated component. The difference in height between these two regions is ~ 1.6 nm. Holes of ~ 5 nm in depth are observed only in the 'island-like' circular hydrocarbon domains, whereas the fluorinated film remains fairly uniform and unbroken. **b**, Friction force measurement (also covering $5 \times 5 \mu\text{m}^2$) done simultaneously with the topography image in **a**. The lateral force image indicates higher friction (brighter contrast) over the fluorinated regions. Highest friction is measured on the silicon surface (at the bottom of the holes in the film). No difference in contrast (friction) is observed between the hydrocarbon layers of different height.

[‡] Also at the IBM Almaden Research Center, San Jose, California 91520, USA.

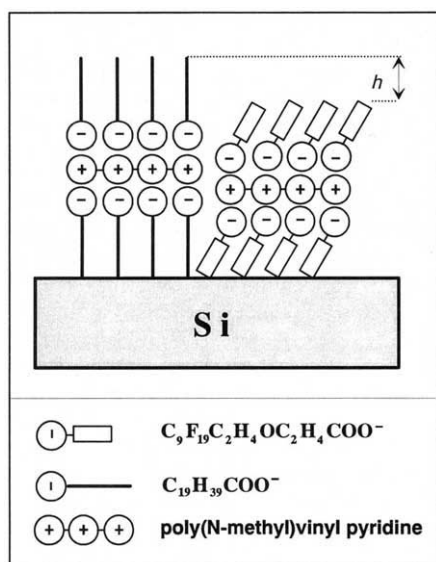


FIG. 2 Schematic view of the LB bilayer system imaged by AFM in Fig. 1. The 'h' represents the height difference between the hydrocarbon and fluorocarbon species (see text).

respectively, we expect a difference in height of ~ 1.0 nm between domains of the two components in this bilayer LB system. Figure 1a shows a step height from flat film to island of 1.6 nm; the slightly greater difference could be due to a greater tilt angle in the fluorocarbon domains²⁷.

In other experiments, AFM images have been recorded on LB bilayer films composed of a single carboxylic acid component, one from arachidate and one from the partially fluorinated carboxylate⁹. The fluorinated acid film contains fewer defects than the arachidic acid film⁹. These results support the assignment of the circular domains to the hydrocarbon component, as holes are observed only in the circular domains in the mixed films. These holes are ~ 5.0 nm in depth, consistent with the thickness of a bilayer. Furthermore, increasing the force while scanning disrupts the hydrocarbon but not the fluorinated areas, in both the one- and two-component films. The fluorinated sites show good resistance to rupture during sliding, in agreement with tribology experiments using macroscopic measurement techniques²³.

Lateral force measurements (Fig. 1b) made simultaneously with the topography measurements (Fig. 1a) indicate roughly four times higher friction over the fluorinated regions than over the hydrocarbon regions. Holes in the film reveal the silicon substrate. The silicon, appearing as dark (low) areas in the topography image, shows a friction force a factor of 10 higher than the hydrocarbon domains in the lateral force image. This difference in friction agrees well with other experiments on both macroscopic and nanometre scales^{28–30}. In sum, from the FFM measurements on these phase-separated LB films, the relative frictions of the hydrocarbon, fluorocarbon and silicon surfaces are shown to be 1:4:10. In the range of nondestructive forces applied in this study, $\sim 10^{-9}$ – 10^{-8} N, no dependence of the friction on load is observed.

The differences in friction are not due to changes in topography. This is demonstrated particularly well by the ability of the friction measurements to identify the scattered islands of 'extraneous' material sitting atop hydrocarbon domains. Whereas normal force measurement determines only the geometry of the island, lateral force measurement determines its composition. The frictional response of the island corresponds to that of the circular hydrocarbon domains, thereby allowing it to be identified as hydrocarbon in nature. This assignment would not be possible with the normal AFM alone.

The friction on the fluorinated areas is higher than on the hydrocarbons. From the performance of fluorine-containing lubricants, such as teflon (PTFE), a reduction in friction might be expected. But it is known from surface force apparatus

experiments²³ that fluorinated LB films have larger shear strengths (friction force divided by the real area of contact in shear) than their hydrogen-containing counterparts. The higher friction measured with the FFM over the fluorocarbon is attributed, at least in part, to the greater stiffness and closer packing³¹ of the fluorocarbon moiety.

These observations of friction and robustness on the scale of nanometres can be applied to the study of tribology, particularly boundary lubrication. Boundary lubrication, as established by Hardy³², deals with the layers of lubricant closest to the solid surfaces undergoing frictional contact, and is of importance in most surface-on-surface sliding mechanisms. The particular advantage of the fluorinated film is its resistance to rupture as known from both the surface force apparatus and AFM measurements. The good performance of fluorocarbons as lubricants can therefore be attributed to high stability towards applied stress, and a reduced friction (compared with unlubricated surfaces) which reduces the wear to very low values (reduction of 10,000)³³.

In nanometre-scale friction and topography measurements, local elasticity plays a role. In our on-going work, we measure a greater elasticity on the molecular scale for the hydrocarbon moieties than the fluorocarbon moieties, evidence for the increased stiffness of fluorinated segments³¹. This is consistent with the higher in-plane cohesive binding that we observe in the fluorocarbon domains, also attributed to the higher barriers for *trans/gauche* isomerization in the stiffer fluorocarbon chains^{34,35}. Even though our observation of lower friction on the hydrocarbon film suggests that hydrocarbon acids could be more effective lubricants than the fluorinated compounds, our comparative measurements show that pure arachidate films and domains are more disrupted by sliding friction than their fluorocarbon counterparts^{9,36}. We propose that a good lubricating mixture would have a fluorinated component to offer good rupture resistance and a hydrocarbon component to contribute low friction, as in the mixture used here. □

Received 6 April; accepted 28 July 1992.

- Knobler, C. J. *Phys.: Condens. Matter* **3**, S17–S22 (1991).
- Möbius, D. & Möhald, H. *Adv. Mater.* **3**, 19–24 (1991).
- Seul, M. & Sammon, M. J. *Phys. Rev. Lett.* **64**, 1903–1906 (1990).
- Chi, L. F., Johnston, R. R. & Ringsdorf, H. *Langmuir* **7**, 2323–2329 (1991).
- Binnig, G., Quate, C. F. & Gerber, C. *Phys. Rev. Lett.* **56**, 930–933 (1986).
- Meyer, E. *et al. Nature* **349**, 398–400 (1991).
- Bourdieu, L., Silberzan, P. & Chatenay, D. *Phys. Rev. Lett.* **67**, 2029–2031 (1991).
- Alves, C., Smith, E. & Porter, M. J. *Am. chem. Soc.* **114**, 1222–1227 (1992).
- Frommer, J. *Angew. Chem.* **31** (in the press).
- Mate, C. M., McClelland, G. M., Erlandsson, R. & Chiang, S. *Phys. Rev. Lett.* **59**, 1942–1945 (1987).
- Meyer, G. & Amer, N. M. *Appl. Phys. Lett.* **57**, 2089–2091 (1990).

12. Marti, O., Colchero, J. & Mlynek, J. *Nanotechnology* **1**, 141–144 (1990).
13. Erlandsson, R. *et al.* *J. Chem. Phys.* **89**, 5190–5193 (1988).
14. Binnig, G., Rohrer, H., Gerber, C. & Weibel, E. *Appl. Phys. Lett.* **40**, 178–181 (1982).
15. Binnig, G. *et al.* *Europhys. Lett.* **3**, 1281–1286 (1987).
16. Meyer, E. *et al.* *J. Vac. Sci. Technol.* **B2**, 1329–1332 (1991).
17. Feenstra, R. M., Stroscio, J. A., Tersoff, J. & Fein, A. P. *Phys. Rev. Lett.* **58**, 1192–1195 (1987).
18. Hamers, R. J., Avouris, Ph. & Bozso, F. *J. Vac. Sci. Technol.* **A6**, 508–511 (1988).
19. Behm, R. J. *Scanning Tunneling Microscopy and Related Methods* (eds Behm, R. J., Garcia, N. & Rohrer, H.) 173–209 (NATO ASI Series, 1990).
20. Feenstra, R. M. *Scanning Tunneling Microscopy and Related Methods* (eds Behm, R. J., Garcia, N. & Rohrer, H.) 211–240 (NATO ASI Series, 1990).
21. Bowden, F. P. & Tabor, D. *The Friction and Wear of Solids* (Oxford Univ. Press, London, 1950).
22. Israelachvili, J. N. & Tabor, D. *Wear* **24**, 386–390 (1973).
23. Briscoe, B. J. & Evans, D. C. B. *Proc. R. Soc. A* **380**, 389–407 (1982).
24. Murray, R. W. *Electroanalytical Chemistry*, Vol. 13 (ed. Bard, A. J.) 191–368 (Dekker, Marcel, New York, 1984).
25. Fujihira, M. *Topics in Organic Electrochemistry* (eds Fry, A. J. & Britton, W. E.) 255–294 (Plenum, New York, 1986).
26. Ullman, A. *An Introduction to Ultrathin Organic Films from Langmuir–Blodgett to Self-Assembly* (Academic, New York, 1991).
27. Naselli, C., Swalen, J. D. & Rabolt, J. F. *J. chem. Phys.* **90**, 3855–3860 (1989).
28. Briscoe, B. & Tabor, D. *Interfacial Phenomena in Apolar Media* (eds Eicke, H.-F. & Parfitt, G. D.) 327–356 (Dekker, Marcel, New York, 1987).
29. Meyer, E. *et al.* *Fundamentals of Friction* (eds Pollock, M. & Singer, I.) (Kluwer, in the press).
30. Meyer, E. *et al.* *Ultramicroscopy* (in the press).
31. Barton, S. W. *et al.* *J. chem. Phys.* **96**, 1343–1351 (1992).
32. Hardy, W. B. *Proc. R. Soc. A* **288**, 313–315 (1913).
33. Rabinowitz, E. & Tabor, D. *Proc. R. Soc. A* **208**, 455–474 (1951).
34. Hsu, S. L. *et al.* *Macromolecules* **23**, 4565 (1990).
35. Bates, T. W. & Stockmayer, W. H. *Macromolecules* **1**, 12 (1968).
36. Meyer, E. *et al.* *Thin Solid Films* (in the press).
37. Umemura, J. *et al.* *Thin Solid Films* **178**, 281 (1989).
38. Nishiyama, K., Kurihara, M. & Fujihira, M. *Thin Solid Films* **179**, 477 (1989).
39. Schreck, M. *et al.* *Thin Solid Films* **175**, 95–105 (1989).

ACKNOWLEDGEMENTS. The work in Switzerland was supported by the Swiss National Science Foundation, the Kommission zur Förderung der wissenschaftlichen Forschung. The work in Japan was supported by the Ministry of Education, Science, and Culture and by the Nissan Science Foundation.

Increased ultraviolet radiation in New Zealand (45° S) relative to Germany (48° N)

G. Seckmeyer*‡ & R. L. McKenzie†‡

* GSF, Institut für Biochemische Pflanzenpathologie, Ingolstädter Landstrasse 1, D-8042 Neuherberg, Germany

† DSIR Physical Sciences, Lauder 9182, Central Otago, New Zealand

RECENT analyses of global ozone measurements have confirmed that ozone reductions are not confined to the Antarctic, but now extend to mid-latitudes in both hemispheres^{1,2}. Ozone reductions lead to increases in biologically damaging ultraviolet radiation³, and such increases have been observed in Antarctica^{1,4} and Australia⁵. Little is known, however, about hemispheric differences in ultraviolet intensities. Here we use a combination of spectral measurements made in Germany and New Zealand with the same spectroradiometer, together with model calculations, to show that in the New Zealand summer of 1990–1991 biologically weighted ultraviolet irradiances were nearly a factor of two greater than those in the summer at similar northern latitudes in Germany. These differences are larger than expected^{3,6}, and are due mainly to decreased stratospheric ozone over New Zealand and increased levels of tropospheric ozone over Germany.

The instrumentation⁷ comprises a horizontal cosine-response diffuser coupled by a quartz fibre optic and an optical chopper to a 0.3-m double monochromator with spectral resolution 1 nm. A photomultiplier detector is used, and data acquisition over the spectral range 290–410 nm is computer-controlled. The scan duration is typically 5 minutes, depending on the light levels. We used the same instrument (temperature-stabilized to 20.0 ± 0.5 °C), method and calibration standards for all measurements, thus avoiding the problem of instrument intercalibration.

Ideally, any comparison should cover an extended period, with continuous observations over a wide range of solar elevations and weather conditions. This has not yet been possible, and here we investigate clear-sky irradiances only. Before the New Zealand campaign, measurements were made at Neuherberg (48.2° N, 11.5° E, altitude 500 m) and the Wank Mountain (47.5° N, 11.1° E, 1,800 m) in Southern Germany in short campaigns since 1986. In New Zealand, data were obtained continuously over daylight hours at Lauder (45° S, 170° E, 300 m) in the summer between mid-November 1990 and mid-March 1991. Data before February were not useful, however, because of calibration difficulties. Good-quality clear-sky data were available only on 21 and 28 February, and 6, 7, 9 and 10 March. Subsequently, data were taken at Neuherberg on 13 July 1990 and on selected days between 26 July and 13 September 1991 (clear skies 29, 30 July). Ozone column measurements from a Dobson spectrophotometer, and profiles measured with balloonsondes, were available at Lauder and Hohenpeissenberg (60 km southwest of Neuherberg). Figure 1 gives an example of the clear-sky ultraviolet irradiances measured at 1 nm spectral resolution and for the same solar zenith angle (34.3°) in Neuherberg and Lauder: the Lauder irradiance at 300 nm (strong ozone absorption) is 2.72 times as great as in Germany, but at 330 nm (weak absorption by ozone) it is only 1.05 times as great. These large differences are mainly due to the difference of 86 Dobson units (DU) in total column ozone. A calculation of differences for average conditions is given later. The spectra were multiplied by two biological weighting functions that represent the relative damaging effect for DNA⁸ and for human skin (erythema)⁹. The weighted irradiance measurements were, respectively, 1.9 and 1.6 times as large at Lauder as at Neuherberg, mainly because of the lower ozone amounts at Lauder.

To investigate whether the irradiance differences were typical for the summer, we compared measured and modelled irradiances. Ozone column measurements were used to calculate ultraviolet spectra, using an adaptation of a simple parametric clear-sky radiative transfer model known as the 'Green model'^{10,11}. Input parameters for the model are: extraterrestrial irradiance, Sun elevation angle, ozone column, height above sea level, albedo, turbidity parameter (aerosol optical depth) and air pressure. The adaptation includes the use of a solar reference spectrum^{12,13} convolved with the instrument slit function. All the clear-sky measurements showed good agreement with calculated spectra (Fig. 2). Although there are wavelength-dependent differences of up to 40%, the biologically weighted irradiances agree to within 5–10%. The measured and calculated irradiances in Neuherberg show similar agreement when ozone values (from Hohenpeissenberg) are reduced by 3%. Results from a recent

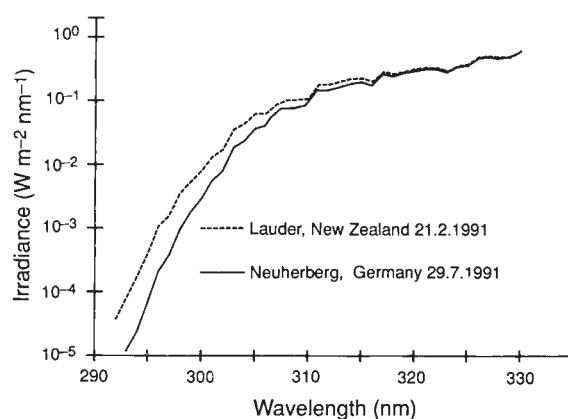


FIG. 1 Measured spectral irradiance in New Zealand and Germany for two summer conditions for solar zenith angle 34.3°. The ozone column was 266 DU in New Zealand and 352 DU in Germany (1 DU = 2.69 × 10¹⁶ molecule cm⁻²).

‡ Address for G.S. from 1 October: Fraunhofer-Institut für Atmosphärische Umweltforschung (IFU), Kreuzackbahnstrasse 19, 8100 Garmisch-Partenkirchen, Germany. Present address for R.L.M.: National Institute of Water and Atmospheric Research Ltd, Te Tari Taihoro Nukurangi (NIWAR), Atmospheric Division, Lauder, Private Bag, Omakau, New Zealand.

recalibration of their Dobson ozone spectrometer imply that a correction of 1–3% may be required (H. Claude, personal communication). Alternatively, it is possible that these ozone measurements are biased by the higher turbidity in Germany. The difference in biologically weighted irradiance is $5 \pm 2\%$ for such a correction¹.

Comparison of modelled and measured spectra confirmed that the larger global ultraviolet irradiances in New Zealand are primarily due to differences in ozone rather than differences in aerosol extinctions. This is consistent with the results of our supporting observations in Germany and New Zealand, which show that although direct-beam visible irradiances are typically 20% higher at Lauder mainly because of less aerosol scattering in New Zealand, total solar irradiances and illuminances are only 10–13% greater. Taking the higher solar elevation angles and the closer Sun–Earth separation into account, only 3–5% of the differences in the visible spectral region are due to aerosol effects. Aerosol extinctions are, however, expected to be larger in the ultraviolet than the visible¹⁴, and aerosol extinctions during these short campaigns may not be typical.

Having established that the model can be used to calculate the biologically weighted ultraviolet irradiance to within 10% of the measurements, we now use it to identify the factors responsible for the larger irradiance measured in New Zealand. In Fig. 3 we show the calculated ratios of clear-sky irradiances at noon for Southern Germany in the summer of 1990 and Lauder, New Zealand, in the following austral summer. The calculation for Germany was mid-July using ozone values from Hohenpeissenberg, whereas the Lauder calculation was for mid-January. The DNA-weighted and erythema-weighted irradiances are larger by factors of 1.81 and 1.44 respectively in New Zealand. An increase by a factor of 1.07 is expected from the closer Sun–Earth separation in January, and a further factor of 1.08 is expected for DNA-weighted irradiance because of differences in latitude, since at noon the solar zenith angle is 3° smaller in Lauder. Differences due to altitude (air pressure) give a factor of 0.98. The deduced increase in New Zealand due to lower aerosol extinction is by a factor of 1.03 ± 0.1 . Recent results from radiative transfer models imply that larger extinctions might be expected in continental Europe¹⁴.

The increase in DNA-weighted ultraviolet due to ozone differences is 1.58 ± 0.10 and the corresponding factor for erythemally weighted irradiance is 1.28 ± 0.10 . These deduced increases agree with previous measurements¹⁵ within the limits of experimental uncertainties. The average difference in total column ozone between the two sites is 53 DU (Fig. 4), but balloonsonde data at each site show that 20–25 DU of this difference arises from the larger tropospheric ozone column above Southern Germany (where 40 DU is typical in summer

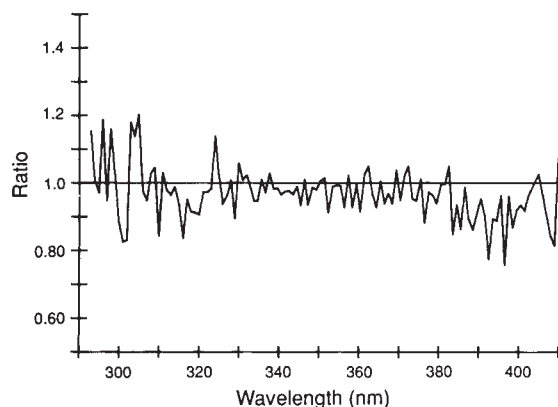


FIG. 2 Example of the ratio of measured to calculated spectral irradiance for the New Zealand data shown in Fig. 1.

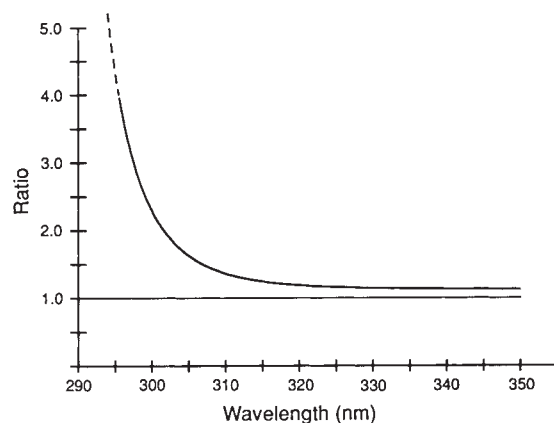


FIG. 3 Calculated ratio of spectral ultraviolet irradiance for summer in Lauder (mid-January) and in Neuherberg (mid-July) at noon, according to the Green model for cloud-free conditions.

for the column below 10 km). Tropospheric ozone is proportionately more efficient per molecule at blocking ultraviolet than stratospheric ozone at these Sun angles¹⁶. But this is a small effect, and we estimate that lower stratospheric ozone over New Zealand and higher tropospheric ozone over Germany have similar contributions to the differences in ultraviolet between the sites.

The differences are much larger than expected from previous studies of hemispheric differences^{3,5}, and this is primarily due to the larger-than-expected difference in ozone. Although the two sites are not precisely conjugate in latitude, gradients in ozone are small in summer at these latitudes. Zonally averaged ozone from the TOMS (Total Ozone Mapping Spectrometer) satellite instrument shows that the 3° difference between 45° and 48° corresponds to an ozone difference <10 DU. The TOMS data also show that over the period 1978–88, trends in summer ozone are only slightly larger at 40 – 50° S than at 40 – 50° N, resulting in ozone columns that are now typically 5–10% (15–30 DU) lower in the south.

To establish whether the differences reported here were typical for the measurement locations in other years, we show, in Fig. 4, time series of ozone measurements made by Dobson spectrometer, and from co-located satellite data. Ozone amounts in New Zealand in January 1991 were lower than usual (except for 1986), whereas in Germany they have been normal for recent

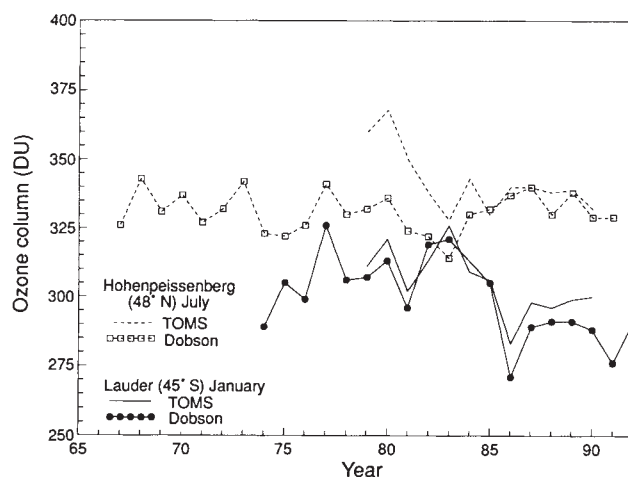


FIG. 4 Time series showing variation of ozone above New Zealand and Germany, measured by Dobson spectrometer, and by the TOMS instrument on the Nimbus-7 satellite.

years. Therefore, the differences in ultraviolet between the sites were larger than normal. In recent years the TOMS and Dobson records at each site show similar variations, although over Lauder the TOMS measurements are ~2–3% higher than the Dobson measurements. The Dobson data show a stronger trend in New Zealand, whereas the TOMS data show a stronger trend in Germany, particularly in the early 1980s. The reasons for the anomalously low ozone in Europe in 1983 have been discussed^{17,18}. Tropospheric ozone levels have increased at Hohenpeissenberg in recent years, and ground based measurements show that the 1980s, the increases in tropospheric ozone have

largely offset the decreases in stratospheric ozone¹⁹. Evidently, the increases in tropospheric ozone have not been fully recognized by the TOMS instrument, so that hemispheric differences in ultraviolet received at the surface are larger than would be expected from an analysis of TOMS data alone.

Hemispheric differences in ultraviolet irradiance therefore seem to have increased in recent years. In future, an increase in ultraviolet is expected in Germany, as tropospheric ozone levels stabilize²⁰. There are considerable day-to-day variability as well as diurnal changes in ultraviolet, which may well be an important issue in future studies. □

Received 25 March; accepted 10 July 1992.

1. *Scientific Assessment of Ozone 1991*, WMO Ozone Rep. No. 25 (eds Watson, R. T. & Albritton, D. L.) (1992).
2. Stolarski, R. S., Bloomfield, P., McPeters, R. D. & Herman, J. R. *Geophys. Res. Lett.* **18**, 1015–1018 (1991).
3. Madronich, S. *Geophys. Res. Lett.* **19**, 37–40 (1992).
4. Lubin, D., Frederick, J. E., Booth, C. R., Lucas, T. & Neuschuler, D. *Geophys. Res. Lett.* **16**, 783–785 (1989).
5. Roy, C. R., Gies, H. P. & Elliot, G. *Nature* **347**, 235–236 (1990).
6. McKenzie, R. L. *Weather. Clim.* **11**, 3–14 (1991).
7. Seckmeyer, G. *Met. Rundsch.* **41**(6), 180–183 (1989).
8. Caldwell, M. M. *Physiological Plant Ecology*, Ch. 6 (Springer, Berlin, 1986).
9. McKinlay, A. F. & Diffey, B. L. *Human Exposure to Ultraviolet Radiation: Risks and Regulations* (eds Passchler, W. R. & Bosnjakovic, B. F.) 83–87 (Elsevier, Amsterdam, 1987).
10. Schipnick, P. F. & Green, A. E. S. *Photochem. Photobiol.* **35**, 89–101 (1982).
11. Rundel, R. *Atmospheric Ozone Reduction and Plant Life* (eds Worrest, R. C. & Caldwell, M. M.) (Springer, Berlin, 1986).

12. Arvesen, J. C., Griffith, R. N. & Pearson, B. D. *Appl. Optics* **8**, 2215–2232 (1969).
13. Nicolet, M. *Planet. Space. Sci.* **29**, 951–974 (1981).
14. Liu, S. C., McKeen, S. A. & Madronich, S. *Geophys. Res. Lett.* **18**, 2265–2268 (1991).
15. McKenzie, R. L., Matthews, W. A. & Johnston, P. V. *Geophys. Res. Lett.* **18**, 2269–2272 (1991).
16. C. Brühl, & Crutzen, P. J. *Geophys. Res. Lett.* **16**, 703–706 (1989).
17. Mantis, H. T., Zerefos, C. S., Bais, A., Ziomias, I. & Kelesiss, A. *Arch. Met. Geoph. Biocl., Ser B* **36**, 135–145 (1986).
18. Bojkov, R. D. *Mon. Weath. Rev.* **115**, 2187–2201 (1987).
19. Wege, K., Claude, H. & Hartmannsgruber, R. *Ozone in the Atmosphere*, Proc. Quadr. Ozone Symp. 1988, (eds Bojkov, R. D. & Fabian, P.) 109–112 (Depak, 1989).
20. Gery, M. W., *UNEP Environmental Effects Panel Rep. 1989*, Ch. 5, 49–54 (1989).

ACKNOWLEDGEMENTS. We thank the German meteorological service (DWD) at Hohenpeissenberg for their assistance. Dobson ozone measurements at Lauder were provided by W. A. Matthews, and TOMS data were provided by R. D. McPeters at NASA Goddard.

Changes in the hydrochemistry of the Black Sea inferred from water density profiles

Suleyman Tugrul, Ozden Basturk, Cemal Saydam & Aysen Yilmaz

Institute of Marine Sciences, Middle East Technical University, PO Box 28, Erdemli 33731, Icel, Turkey

DURING the past two decades, catastrophic changes have occurred in the Black Sea ecosystem: the influx of pollution from the major rivers has caused intense eutrophication at the northwest coastal margin¹, and fish stocks have collapsed throughout the sea². The hydrochemical details of these events are still poorly understood^{3–7}, and a way needs to be found to distinguish long-term variations from short-term natural fluctuations^{3,4} if future management of the Black Sea ecosystem is to be successful. We show here that a coherent description may be achieved by analysing the hydrochemical data as a function of water density rather than depth. Our findings suggest that, contrary to the suggestion of Murray *et al.*³, the upper boundary of the low-lying anoxic waters has remained stationary since 1969, whereas the intermediate suboxic zone has enlarged, reducing the overall depth of the oxygenated upper waters by ~20 m. Moreover, a long-term increase in the nitrate concentration and a concomitant decrease in the silicate and ammonia concentrations in this upper layer are indicative of the considerable changes taking place in the biochemical regime of the Black Sea.

Put simply, the Black Sea, an enclosed sea and the largest anoxic basin in the world, consists of a deep basin of sulphidic saline water overlain by ~100 m of oxygenated brackish water^{8,9}. Between these two layers, an oxic–anoxic interface of a few tens of metres has developed across which the magnitude of the electrochemical oxidation–reduction (redox) potential changes markedly^{8,9}. This suboxic (chemocline) zone coincides with the vertically stable layer^{9,10} (pycnocline) in which the water density increases in parallel with the salinity. The depth of the pycnocline and of the associated suboxic hydrochemistry varies both

seasonally and from one part of the Black Sea to another depending on the circulation and the intensity of eddies^{7–11}. Comparison of depth profiles at limited locations is therefore unlikely to define long-term changes in the hydrochemistry of the fertile waters of the Black Sea. But diffusion and mixing in the Black Sea is rapid along equal-density (isopycnal) surfaces and much slower in the vertical direction^{10,11}. Consequently, chemical concentrations in the suboxic zone and the upper anoxic layer plotted against density, rather than depth, may be expected to give individually similar density profiles, independent of seasons and location. Codispoti *et al.*⁷ demonstrated this for nitrate concentrations measured in 1988. Here we extend their observations to include the dissolved oxygen, hydrogen sulphide and nutrient data collected during the 1969 *Atlantis II*¹², 1988 *Knorr*¹³ and 1991 *Bilim* (our own unpublished data) research cruises throughout the Black Sea or within the Turkish Exclusive Economic Zone.

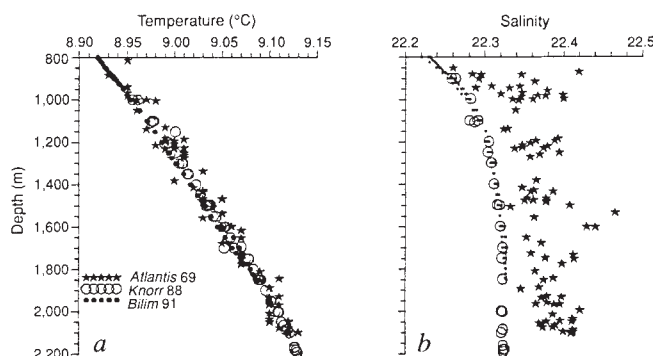


FIG. 1 Basin-wide temperature and salinity profiles in the anoxic waters of the Black Sea. *a*, Temperature and *b*, salinity plotted against water depth to 2,000 m in 1969, 1988 and 1991. We include all the temperature and salinity data from the 1969 *Atlantis II*¹² and 1988 *Knorr*¹³ cruises, and data from selected stations of the September 1991 *Bilim* cruise, in the composite profiles. Recent temperature and salinity data were measured *in situ* with a SeaBird model CTD (conductivity, temperature and depth) instrument. In 1969, temperature and salinity were determined by reversing thermometers attached to bottles and by a Hytech inductive salinometer¹².

In plotting data from different cruises it has been necessary to check the consistency of the density (σ_t) values. Our test for consistency has been the assumption that at depths below 1,000 m, temperatures and salinities have remained constant during recent decades. Figure 1 shows this to have been true to within $\pm 0.01^\circ\text{C}$ of temperatures measured since 1969. Salinities for 1969, however, were always higher than recent data by an average of 0.065 ± 0.02 (33 measurements below 1,700 m). Corrected σ_t values (obtained by using measured temperatures, subtracting 0.065 from the observed salinities and using an equation of state¹⁴ which, in itself, reduces σ_t values by 0.05 units) are, in total, about 0.1 σ_t units smaller than the values reported by Brewer¹². Such a difference, not previously noticed^{3,7}, produces an artificial discrepancy between density profiles from the uncorrected data of 1969 and recent observations.

Previous workers^{3,5,7-9} have naturally attempted to define the upper and lower boundaries of the oxic-anoxic interface by examining the depth profiles of dissolved oxygen (O_2) and hydrogen sulphide (H_2S), whose concentrations in this zone are low and can be measured accurately only by special techniques^{5,7}. From changes observed in depth profiles at a single station in the central Black Sea, Murray *et al.*⁴ suggested that the anoxic interface had risen by 30 m since 1969, whereas basin-wide comparison of the depths of the anoxic interface seemed to indicate that changes in this boundary had been insignificant in recent decades⁵. Here we discuss this problem using composite profiles of sulphide plotted against density (Fig. 2). Hydrogen sulphide concentrations from the high-resolution data of the 1988 *Knorr* cruise¹³ increase smoothly with density, indicating that similar sulphide concentrations follow similar density surfaces throughout the basin (Fig. 2*b*). A concentration of $[\text{H}_2\text{S}] = 10 \mu\text{M}$ was observed consistently at $\sigma_t = 16.32$. The composite profiles from the 1969 and 1991 data are much scattered at each given density surface (Fig. 2*a, c*), strongly suggesting that sampling and analytical techniques were inadequate. The profile for 1991 is relatively consistent with that of 1988, whereas the 1969 profile coincides with recent profiles only at larger sulphide concentrations below the $\sigma_t = 16.4$ surface (or 30 m from the sulphidic boundary). A hydrogen sulphide concentration of $10 \mu\text{M}$ was not reached in 1969 until $\sigma_t = 16.35$ – 16.40 , but the large scatter in the 1969 profile, by as much as $20 \mu\text{M}$ at a given density surface, strongly suggests that the detection of sulphide at larger values of σ_t was inaccurate by the equivalent of more than 0.05 σ_t units (or 10 m in depth) probably because of the incomplete recovery of sulphides from samples that were stored until analysis¹².

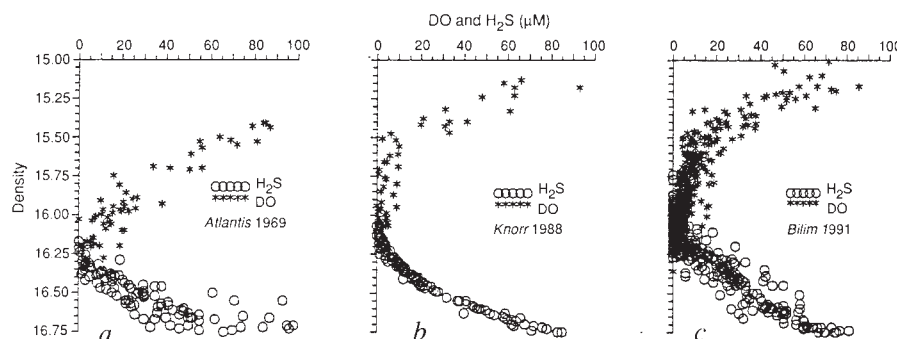
To clarify this problem further we have examined the phosphate concentrations, whose profile within the anoxic transition (redox) zone is dominated by redox processes^{7,15,16}. The phosphate profiles show two maxima^{7-9,15,16}; the upper, broad

maximum occurs at similar depths to the nitrate maximum whereas the deeper, larger maximum is formed at depths within the anoxic transition zone at which the sulphide concentrations first exceed $3 \mu\text{M}$. The composite profiles of phosphate shown in Fig. 3*a* and *b* demonstrate that the phosphate concentrations in the suboxic-anoxic transition zone increase by at least an order of magnitude from a minimum value at $\sigma_t = 15.9$ – 16.0 to a maximum at $\sigma_t = 16.20$ – 16.25 . The location of the sharp increase in the concentrations coincides with similar features observed in the Fe^{2+} and Mn^{2+} profiles^{17,18}, suggesting that redox processes involving iron and manganese are important in the vertical transport of phosphate and in the formation of the deeper phosphate maximum^{15,16}. The σ_t values corresponding to these characteristic features and which coincide with those locating the suboxic-sulphidic interface^{7-9,15,16} have remained constant since 1969, so the redox potentials defining the anoxic interface have remained at this chemical depth since 1969.

The hydrochemistry of the fertile upper layer which all the living resources inhabit is more likely to be affected by recent changes in the Black Sea ecosystem than is the anoxic lower layer. Murray *et al.*⁴, comparing the depth profiles of oxygen from the central Black Sea, suggested that the oxygenated upper layer has become shallower because of the formation of a relatively thick (few 10 m), oxygen-deficient zone (termed here the suboxic zone, where $[\text{O}_2] < 10 \mu\text{M}$ and $[\text{H}_2\text{S}] < 3 \mu\text{M}$) at the oxic-anoxic interface. The oxygen profiles in Fig. 3 show that discussions of the development of the suboxic zone are hampered by the scatter in the data. Nevertheless, in 1969, oxygen concentrations throughout the Black Sea exceeded $10 \mu\text{M}$ down to at least the $\sigma_t = 16.0$ surface, whereas in 1988 and in 1991, oxygen decreased to suboxic concentrations ($< 10 \mu\text{M}$) by the $\sigma_t = 15.6$ – 15.7 surfaces in the Turkish Exclusive Economic Zone. In recent years, the suboxic zone therefore seems to have spread towards the surface by $\sim 0.3 \sigma_t$ units relative to its position in 1969 or by at least 20 m with respect to the $\sigma_t = 16.0$ surface. Confirmation that the shift has been caused by long-term changes in the system is provided by a similar upward shift in the depth of the nitrate maximum, formed as the result of nitrification and denitrification⁷⁻⁹, at depths where oxygen concentrations diminished below $10 \mu\text{M}$ at all seasons throughout the Black Sea. The nitrate profiles shown in Fig. 3*c* and *d* demonstrate that the maximum concentrations have remained stationary at $\sigma_t = 15.3$ – 15.5 since 1988, whereas in 1969, the maximum was established between the $\sigma_t = 15.7$ and 15.9 surfaces.

Nitrate concentrations (Fig. 3*c* and *d*) in the oxic layer, which is thus becoming shallower, have increased at least two- to threefold from concentrations of 2–4 μM in the 1960s (refs 8, 9, 19) to 7–9 μM by 1991 (refs 7, 20). The ratio of nitrogen in the form of $\text{NH}_3 + \text{NO}_3$ to phosphorus in the form of PO_4 in the oxic layer between the $\sigma_t = 15.0$ and 15.7 surfaces shows

FIG. 2 Composite profiles of dissolved oxygen (DO) and hydrogen sulphide in the oxic-anoxic transition zone of the Black Sea relative to density (σ_t). *a*, Historical profiles of dissolved oxygen and hydrogen sulphide from 30 stations of the March–April 1969 *Atlantis II* Black Sea cruise. *b*, Recent composite profiles from five oxygen and 12 sulphide stations of the June–July 1988 *Knorr* cruise in the Turkish Exclusive Economic Zone. *c*, Latest profiles from 46 stations of the September 1991 *Bilim* cruise in this zone. Oxygen was determined by Winkler titration^{12,13,20} except for the 1988 *Knorr* data below $20 \mu\text{M}$, which were obtained by colorimetry⁷. Sulphide data for 1969 were determined manually by colorimetry in samples from bottle casts, stored before analysis by adding preservative¹². The high-resolution sulphide data for 1988 were produced by a multichannel autoanalyser coupled to a continuous pumping system^{7,13}. The latest data



from the 1991 *Bilim* cruise were measured by iodimetry²⁰ in the subsamples from a 12-bottle Rozette attached to a CTD.

that the N/P ratio beneath the euphotic zone (where only 0.1% of the surface light penetrates), which is lower than the Redfield value for the oceans (N/P = 16 by mole), has varied only between 6 and 8 since the 1960s (refs 7, 9, 21). The ratio, dominated by both ammonia and nitrate until 1975, has become nitrate-dependent since 1980. But the N/P ratio evident from the *Atlantis II* 1969 data was as low as 2–3 because of the underestimation of ammonia²² (Fig. 3e). This figure also shows that there has been a net loss of inorganic nitrogen in the bottom of the suboxic zone in recent years. Silicate concentrations in the surface waters of the Black Sea have decreased by an order of magnitude since 1969 (refs 6–9, 20), with a lesser decrease in the concentrations in the lower oxic layer extending down to the $\sigma_t = 15.7$ surface

(Fig. 3f). The present concentrations, as low as 0.3 μM in summer²¹, may be limiting the growth of diatoms. Alterations in the ecosystem of the Black Sea have therefore been observable since the 1970s with concurrent changes in the biochemical processes dominating nitrogen and silicate cycles in the upper layer. □

Received 27 January; accepted 28 July 1992.

1. Bologna, A. S. *Thalassia Yugoslavia* 21/22, 1–22 (1985/1986).
2. Caddy, J. F. & Griffiths, R. C. in *General Fisheries Council for the Mediterranean Studies and Reviews* 63, 43–71 (FAO, Rome, 1990).
3. Murray, J. W. et al. *Nature* 338, 411–413 (1989).
4. Kempe, S., Liebeck, G. & Diercks, A. R. *Nature* 346, 419 (1990).
5. Bezborodov, A. A. in *Complex Oceanographic Investigation of the Black Sea* (in Russian) (ed. Zaika, V. E.) 76–114 (USSR Academy of Sciences, Sevastopol, 1990).
6. Trotsenko, B. G. *Oceanology* 29, 455–456 (1989).
7. Codispoti, L. A., Friedrich, G. E., Murray, J. W. & Sakamoto, C. *Deep-Sea Res.* 38, S691–S710 (1991).
8. Grasshoff, K. in *Chemical Oceanography*, Vol. 2 (eds Riley, J. P. & Skirrow, G.) 545–574 (Academic, London, 1975).
9. Sorokin, Yu. I. in *Estuarine and Enclosed Seas* (ed. Ketchum, B. H.) 253–292 (Elsevier, Amsterdam, 1983).
10. Murray, J. W., Top, Z. & Ozsoy, E. *Deep-Sea Res.* 38, S663–S689 (1991).
11. Oguz, T., Lavolette, P. E. & Unluata, U. *J. geophys. Res.* (in the press).
12. Brewer, P. G. *Woods Hole Oceanogr. Inst. Tech. Note* 75–66 (1971).
13. Friedrich, G. E., Codispoti, L. A. & Sakamoto, C. *Monterey Aquarium Res. Inst. Tech. Rep. No.* 90-3 (1990).
14. Fofonoff, N. P. & Millard, R. C. Jr. *Unesco Tech. Pap. mar. Sci.* 44, 53 (Unesco, Paris, 1983).
15. Shaffer, G. *Nature* 321, 515–517 (1986).
16. Gusarova, A. N. & Sapozhnikov, V. V. *Oceanology* 25, 744–746 (1985).
17. Spencer, D. W. & Brewer, P. G. *J. geophys. Res.* 76, 5877–5892 (1971).
18. Lewis, B. L. & Landing, W. M. *Deep-Sea Res.* 38, S773–S803 (1991).
19. Sen Gupta, R. *Deep-Sea Res.* 75, 457–475 (1971).
20. Oguz, T., Turgrul, S., Bingel, F. & Unsal, M. *METU-Mar. Sci. Inst. Tech. Rep. No.* 90-5 (1990).
21. Kirikova, M. V. *Mar. Ecol.* 23, 3–10 (in Russian, 1986).
22. Brewer, P. G. & Murray, J. W. *Deep-Sea Res.* 20, 803–818 (1973).

ACKNOWLEDGEMENTS. We thank A. Gaines, E. Ozsoy and J. Murray for constructive criticism, E. Hatipoglu, K. Yilmaz and Z. Karabulut for the nutrient, oxygen and hydrogen sulphide measurements, and the crew of the R/V *Bilim* for collaboration during the cruises. This work was funded by the Turkish Scientific and Technical Research Council, TUBITAK, and by the Scientific Affairs Division of NATO within the framework of the Science for Stability Programme.

Intense hydrolytic enzyme activity on marine aggregates and implications for rapid particle dissolution

David C. Smith*, Meinhard Simon†, Alice L. Aldredge‡ & Farooq Azam*

* Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego 0202, La Jolla, California 92093, USA

† Limnological Institute, University of Konstanz, D-7750 Konstanz, Germany

‡ Department of Biological Sciences, University of California, Santa Barbara, Santa Barbara, California 93106, USA

LARGE, rapidly sinking organic aggregates are an important component of the carbon flux from the ocean's surface to its depths. Marine snow, the main type of large (>0.5 mm) aggregate, is heavily colonized by bacteria in surface waters¹, yet the carbon demand of the attached bacteria is so small that months to years are required to consume the aggregates' carbon^{2–5}. This has led to the conclusion that marine aggregates are resistant to degradation by attached bacteria, and thus act as refractory carriers of carbon to the deep ocean. Here we report that aggregates play host to intense activities of hydrolytic enzymes (presumably due to cell surface bound and released enzymes of the attached bacteria), which render the aggregates soluble. Particulate amino acids were hydrolysed rapidly (turnover time 0.2–2.1 days), with very little of the hydrolysate being taken up by the attached bacteria. Our results support the hypothesis^{6,7} that such 'uncoupled' hydrolysis is a biochemical mechanism for large-scale transfer of organic matter from sinking particles to the dissolved phase,

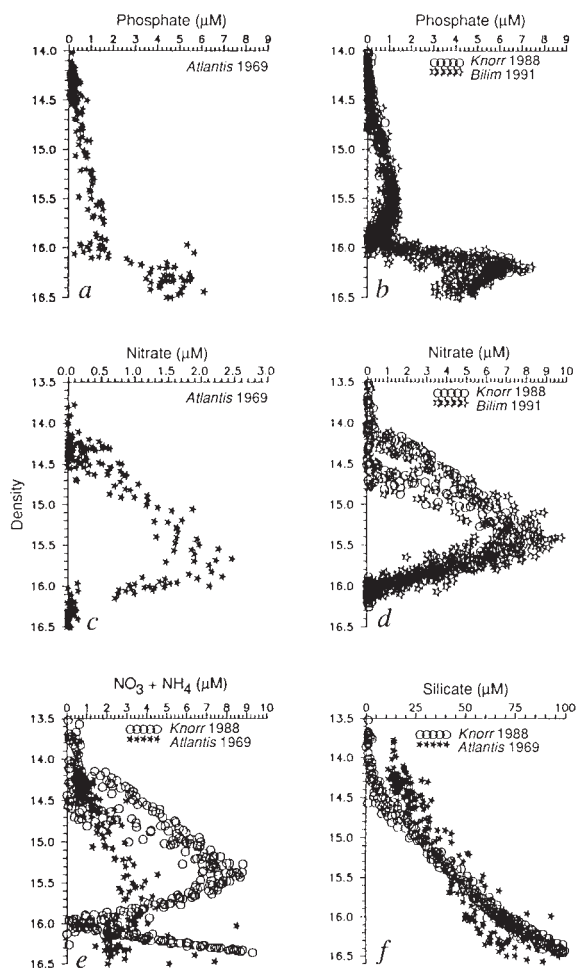


FIG. 3 Basin-wide vertical distribution of nutrients in the upper layer of the Black Sea relative to density. Composite profiles of (a, b) phosphate (c, d) nitrate in 1969, 1988 and 1991 plotted against σ_t using the data sets from 27 stations of the 1969 *Atlantis II*, 12 stations of the 1988 *Knorr* and 46 stations of the 1991 *Bilim* Black Sea cruises. e, Composite profiles of nitrate plus ammonia for 1969 and 1988, showing an inorganic nitrogen deficiency in the lower suboxic zone resulting from denitrification as well as a linear increase in ammonia concentrations with water density in the upper anoxic lower layer. The 1969 data alone show considerable scatter, probably because of analytical artefacts²¹. f, Silicate profiles for 1969 and 1988 plotted from the data sets of the same nutrient stations, showing that in 1969 the silicate concentrations in the upper layer down to the suboxic zone were larger than recent values, although values for deep anoxic water appear¹² underestimated by ~30% when compared with those obtained by others^{9,13}. In e and f, no data are given for 1991 because of a lack of ammonia and silicate measurements. The nutrient samples of 1969 and 1991 collected by discrete bottle casts were stored before being autoanalysed by colorimetry^{12,20}, whereas the 1988 samples were processed immediately by a multichannel autoanalyser coupled to a pump-profiling system^{7,14}.

TABLE 1 Aggregate composition, turnover and release of dissolved organic matter

Date in April 1990	Aggregate type	Carbon content ($\mu\text{g C agg}^{-1}$)	Bacterial C demand ($\text{ng C agg}^{-1} \text{ d}^{-1}$)	C turnover time (d)	PCAA (ng agg^{-1})	DCAA release ($\text{ng agg}^{-1} \text{ d}^{-1}$)
19	Larvacean house	3.5	8.69	403	200	936
20	Diatom floc	3.5	6.89	508	1126	527
21	Faecal pellet	1.5	1.05	1429	309	ND
23	Larvacean house	3.2	5.56	576	226	478
25	Larvacean house	4.5	10.50	429	571	365

Date in April 1990	Aggregate type	PCAA turnover T_h (d)	T_b (d)	Release uptake %	Mixed layer (m)	Total agg. vol. ($\mu\text{l l}^{-1}$)	Flux from aggregates DCAA ($\text{mg m}^{-2} \text{ d}^{-1}$)	DCAA (N)
19	Larvacean house	0.2	11.5	98.2	20	116.1	155.2	23.3
20	Diatom floc	2.1	81.7	97.5	27	15.4	15.4	2.4
21	Faecal pellet			50	367.6			
23	Larvacean house	0.5	20.3	97.7	35	44.9	53.6	8.0
25	Larvacean house	1.6	27.2	94.6	17	6.9	3.1	0.5

Carbon content of the aggregates was determined with a CHN analyser (Leeman Labs, Model 440). Bacterial carbon demand (bacteria carbon production + respiration) was determined on aggregates using the [^3H]leucine incorporation method^{13,14} and assumed 30% growth efficiency²⁷. Particulate combined amino acids (PCAA) were measured in aggregates as described¹³. Aggregates were incubated in 0.2- μm filtered sea water in the dark at surface sea water temperature ($16 \pm 1^\circ\text{C}$) and subsamples of surrounding sea water were collected with time and filtered for subsequent DCAA analysis as described¹³. PCAA turnover times were calculated using the DCAA release rate and bacterial carbon demand. Release/uptake calculated as DCAA release/DCAA release + bacterial carbon demand. Abbreviations: PCAA, particulate combined amino acids; DCAA, dissolved combined amino acids; T_h , carbon turnover time due to hydrolysis; T_b , carbon turnover time due to bacterial carbon demand; ND, not determined; agg, aggregate.

and may supply the slowly degradable dissolved organic matter for downward export postulated by recent models⁸⁻¹⁰.

Individual aggregates were collected by SCUBA in the upper 25 m during two cruises on the RV 'Point Sur' (in April 1990 and 1991) in the Southern California Bight ($33^\circ 47' \text{N}$, $119^\circ 03' \text{W}$). An *in situ* camera system¹¹ recorded video and still images which were digitized to calculate aggregate abundance and volume. Using a highly sensitive fluorogenic assay¹², we quantified the activities of several hydrolytic enzymes on individual aggregates. Bacterial carbon demand on aggregates and in sea water was determined by ^3H -leucine incorporation^{13,14}. Particulate (PCAA) and dissolved ($<0.2 \mu\text{m}$; DCAA) combined amino acids were measured by high-performance liquid chromatography after acid hydrolysis to yield individual amino acids¹⁴.

The carbon demand (BCD) of attached bacteria required $>1 \text{ yr}$ to use up the aggregates' carbon (Table 1), despite the intense hydrolytic enzyme activity of the aggregates, which was presumably due to high concentrations of bacteria. We presumed that enzyme hydrolysis of the aggregates is rapid, but, given the low carbon demand of the attached bacteria, most of the hydrolysate diffuses away. To test this, we measured the PCAA content of the aggregates and the rate of DCAA release into the surrounding sea water. PCAA turnover times due to DCAA release ($T_h = \text{PCAA/DCAA d}^{-1}$) were very rapid ($1.1 \pm 0.9 \text{ d}$; range, 0.2–2.1 d). In contrast, PCAA turnover times resulting from bacterial carbon demand ($T_b = \text{PCAA/BCD d}^{-1}$) were much longer ($\geq 35.2 (\pm 31.7) \text{ d}$). These are minimum estimates as we assume all BCD is supported by PCAA. Even these minimum estimates of T_b were 1–2 orders of magnitude higher than T_h , resulting in $\geq 97\% (\pm 1.6)$ of PCAA hydrolysate being released into the surrounding sea water. Protozoan grazing on attached bacteria¹⁵ could not have contributed much to DCAA release because the bacterial protein production was $<0.02\%$ of DCAA release.

Calculations based on data in Table 1 and assuming a sinking rate of $50\text{--}100 \text{ m d}^{-1}$ (ref. 1) indicate that aggregates sink below the mixed layer ($\sim 32 \text{ m}$) in 0.3–0.6 d, in which time 15–320% of the aggregates' PCAA is released. We calculated PCAA solubilization of all aggregates in the mixed layer at our site from the volume abundance of aggregates larger than 0.5 mm by assuming that their solubilization rate equalled that of our aggregates collected by SCUBA (over 2 mm in size). Integrated

DCAA release for the mixed layer was $3.1\text{--}155.2 \text{ mg m}^{-2} \text{ d}^{-1}$ ($8.5 \pm 10.3 \text{ mg N m}^{-2} \text{ d}^{-1}$). Nitrogen solubilization, when compared with vertical flux of nitrogen ($16.7 \text{ mg N m}^{-2} \text{ d}^{-1}$; assuming primary production of $0.390 \text{ g C m}^{-2} \text{ d}^{-1}$, an *f*-ratio of 0.3, and $\text{C/N} = 7$ for this area¹⁶) indicates solubilization was equivalent to 51.2% (3–139%) of the nitrogen sinking flux (that is, nitrogen loss from the mixed layer would be 51.2% greater in the absence of solubilization). This is a minimum estimate because the sinking rates of small particles are $<100 \text{ m d}^{-1}$. High variability in solubilization suggests it could be a significant variable in the sinking flux of nitrogen.

We did not determine whether other aggregate constituents were also rapidly hydrolysed and released, but this is suggested by the high volume concentration factors (VCF; see Table 2 legend) of all the enzymes assayed (Table 2). Sinking particles lose nitrogen faster than carbon, thus increasing their C/N ratio as the depth increases^{17,18}, but the mechanism is unknown. Possibly the nitrogen-rich PCAA (which accounted for 15% of the entire nitrogen pool of the aggregates) is hydrolysed faster than carbon-rich pools such as polysaccharides. This is consistent with sediment trap data showing that amino acids decline rapidly with depth¹⁹. If enzymatic solubilization is the main fate of the total organic matter in aggregates, and the aggregates' C/N ratio increases with depth, it follows that carbon must be solubilized more slowly than nitrogen. Protease activity in the aggregates was $10\text{--}10^3$ times higher than the activity of polysaccharidases, α - and β -glucosidase. An earlier study of mineral-based marine agglomerates also showed that protease activity was higher than β -glucosidase activity²⁰. If the relative hydrolysis rates of fluorogenic substrates measure the relative activities of protease and glucosidases, then our results indicate that protein is solubilized faster than polysaccharide. Such differential solubilization of the aggregates' macromolecular constituents provides a biochemical mechanism for preferential loss of carbon relative to nitrogen from the mixed layer.

Bacteria on aggregates were larger and more concentrated ($1.9 \pm 1.3 \times 10^8 \text{ ml}^{-1}$; $>10^2$ times seawater concentration) and were probably the source of hydrolytic enzymes on the aggregates. Epifluorescence microscopy of aggregates revealed bacteria as the dominant biomass. Larger²¹ and attached²² bacteria reportedly have higher cell-specific ectoprotease activity. Assuming all protease activity was bacterial, the mean cell-specific activity on aggregates ($242.1 \pm 492.5 \text{ attomol cell}^{-1} \text{ h}^{-1}$;

TABLE 2 Hydrolytic enzyme activities and volume concentration factors

Enzyme	18 April 1990			20 April 1990			21 April 1990		
	Agg.	Sea water	VCF	Agg.	Sea water	VCF	Agg.	Sea water	VCF
Protease	887.2 ± 146.9	3.1 ± 0.1	286	1,027.0 ± 913.4	0.40 ± 0.29	2,542	4.9 ± 3.7	0.08 ± 0.01	64
Phosphatase	3,967.6 ± 3,857.2	1.6 ± 0.2	2,527	583.4 ± 318.1	0.01 ± 0.02	48,615	9.9	<0.01 ± 0.01	2,473
β-Glucosidase	17.6 ± 9.3	<0.1 ± 0.0	588	24.8 ± 13.7	0.01 ± 0.00	2,751	<0.01 ± 0.00	0.01 ± 0.00	0
Lysozyme	ND	ND	ND	ND	ND	ND	ND	ND	ND
Chitinase	32.2 ± 11.4	0	∞	22.7 ± 13.8	0.01 ± 0.00	2,524	1.3	0.01 ± 0.00	126
α-Glucosidase	3.1	0	∞	3.7 ± 1.2	0.01 ± 0.01	526	0.5 ± 0.2	0	∞
Chitobiase	ND	ND	ND	ND	ND	ND	ND	ND	ND

Enzyme	23 April 1990			26 April 1990		
	Agg.	Sea water	VCF	Agg.	Sea water	VCF
Protease	163.4 ± 142.0	0.15 ± 0.08	1,089	3,324.1 ± 1,895.0	0.41 ± 0.21	8,127
Phosphatase	265.0 ± 136.1	0	∞	1,638.6 ± 543.8	1.21 ± 0.28	1,356
β-Glucosidase	ND	ND	ND	129.7 ± 77.6	0.01 ± 0.01	10,811
Lysozyme	27.1 ± 10.6	<0.01 ± 0.00	27,072	47.5	ND	ND
Chitinase	3.9 ± 3.2	0.01 ± 0.01	279	ND	ND	ND
α-Glucosidase	ND	ND	ND	ND	ND	ND
Chitobiase	17.1 ± 24.0	0	∞	ND	ND	ND

Enzyme activities on individual aggregates in 1.5 ml sea water were determined using the fluorogenic substrates L-leucine-7- amino-4-methylcoumarin (for protease enzymes), 4-methylumbelliferyl (MUF)-phosphate (for phosphatases), MUF-β-D-glucoside (for β-glucosidase), MUF-α-D-glucoside (for α-glucosidase), MUF-β-D-N, N', N''-triacylchitotriose (for lysozyme), MUF-N-acetyl-β-D-glucosamine (for chitinase) and MUF-β-D-N,N'-diacylchitobioside (for chitobiosidase). The substrates were added to 20 nM final concentration and the aggregates incubated *in situ* at (16 ± 1 °C) in the dark. Assays were done in triplicate; one aggregate for each substrate was heated to 80 °C for 20 min before addition of substrate as a control. Fluorescence was determined with a Farrand System 3 spectrofluorometer or a Hoefer TKO-100 fluorometer calibrated with 7-amino-4-methylcoumarin and 4-methylumbelliferone. Bulk sea water samples were treated in the same way as aggregate samples. All activities are expressed in μmol ml⁻¹ h⁻¹ (±1 s.d.). ND, activity was not determined; zero indicates activity was undetectable. The volume concentration factors (VCF) were calculated as (activity per ml aggregate) – (activity per ml sea water)/(activity per ml sea water).

$n = 13$) was higher and more variable than in sea water samples (52.5 ± 14.8 attomol cell⁻¹ h⁻¹), although this was not statistically significant owing to the high variability. The high variability in per-cell protease activity may reflect species-specific differences of bacteria on aggregates, and/or the chemical composition of the aggregates. Bacteria may use protease as a defence against bacteriophage attack²³ and cell-specific protease activity may also reflect phage activity on aggregates.

As simulation models driven by dissolved organic matter (DOM) better predict the observed oceanic distribution of organic matter than models driven by particulate organic matter (POM)^{9,10}, a large fraction of downward flux of organic matter may occur as slowly degradable DOM⁸ rather than as sinking particles. How the slowly degradable DOM is produced in the upper mixed layer is an important unanswered question. We propose that aggregates, permeated with hydrolases, are sites of production of some of this DOM. Aggregates of biogenic particles are probably dominated by complex mixtures of structural biopolymers (protein, mucopolysaccharides, chitin and cell wall polymers for example) and their hydrolysis would generate high local concentrations of a chemically diverse mixture of dissolved polymers, oligomers and monomers in the matrix of the aggregates. Hydrolysis of structural suprapolymers may produce some slowly degradable DOM. Further, the aggregate's microenvironment may support condensation reactions such as the glucosylation of proteins to produce slowly degradable DOM (glucosylated proteins degrade slowly in sea water; R. G. Keil, personal communication). Whether and to what extent such chemical transformations occur needs to be investigated because they may provide the exportable DOM needed to explain the observed patterns of carbon distribution. Indeed, uncoupled solubilization is consistent with the production of slowly degradable DOM on aggregates.

The enzyme-mediated transformation may not be limited to large aggregates or to the production of true DOM. Sinking aggregates capture small suspended particles from sea water²⁴, which would expose these 'new' particles to intense enzyme

activity in the aggregate. In this hypothetical situation, the aggregate acts as an 'enzymatic reactor', fuelled not only by POM of the aggregate itself, but also by the suspended particles and colloids it accretes. This would result in the gradual accumulation of refractory material, thus prolonging the persistence of the aggregate. Further, the hydrolytic attack on particles, for example by endohydrolases, may release 'sub-micrometre' particles²⁵. In our experiments the aggregates released mostly DCAA rather than DFAA (we did not determine the molecular size distribution of DCAA, however). The abundance of these apparently biogenic sub-micrometre particles (accounting for ≥10% of dissolved organic matter) closely covaries with larger POC (ref. 25). This is consistent with our suggestion that large POC is a source of sub-micrometre particles.

DOM dynamics are of key importance to understanding carbon fluxes in the ocean²⁶, hence we need to elucidate mechanisms that regulate the flux of POM to DOM. We conclude that 'uncoupled' solubilization is an important mechanism by which bacteria exert biochemical pressure on the particulate phase to mediate large-scale transition of POM to DOM. □

Received 5 May; accepted 21 July 1992.

1. Alldredge, A. L. & Gotschalk, C. C. *Cont. Shelf Res.* **10**, 41–58 (1990).
2. Ducklow, H. W., Kirchman, D. L. & Rowe, G. T. *Appl. envir. Microbiol.* **43**, 769–776 (1982).
3. Alldredge, A. & Youngbluth, M. *Deep Sea Res.* **32**, 1445–1456 (1985).
4. Karl, D. M., Knauer, G. A. & Martin, J. H. *Nature* **332**, 438–441 (1988).
5. Simon, M., Alldredge, A. L. & Azam, F. *Mar. Ecol. Prog. Ser.* **65**, 205–211 (1990).
6. Azam, F. *The Radiocological Role of Marine Bacterioplankton 1-2-7* (Institut Royal des Sciences Naturelles de Belgique, Brussels, 1984).
7. Cho, B. C. & Azam, F. *Nature* **332**, 441–443 (1988).
8. Sarmiento, J. L., Toggweiler, J. R. & Najjar, R. *Phil. Trans. R. Soc. Lond.* **325**, 3–21 (1988).
9. Bacastow, R. & Maier-Reimer, E. *Global biogeochem. Cycles* **5**, 71–85 (1991).
10. Najjar, R. G., Sarmiento, J. L. & Toggweiler, J. R. *Global biogeochem. Cycles* **6**, 45–76 (1992).
11. Honjo, S., Doherty, K. W., Agrawal, Y. C. & Asper, V. L. *Deep Sea Res.* **31**, 67–76 (1984).
12. Hoppe, H.-G. *Mar. Ecol. Prog. Ser.* **11**, 299–308 (1983).
13. Kirchman, D., K'Neess, E. & Hodson, R. *Appl. envir. Microbiol.* **49**, 599–607 (1985).
14. Simon, M. & Azam, F. *Mar. Ecol. Prog. Ser.* **51**, 201–213 (1989).
15. Caron, D. A. *Microb. Ecol.* **13**, 203–218 (1987).
16. Eppley, R. W. & Holm-Hansen, O. in *Plankton Dynamics of the Southern California Bight* (ed. Eppley, R. W.) 176–215 (Springer, Berlin, 1986).
17. Suess, E. *Nature* **288**, 260–263 (1980).

18. Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. W. *Deep Sea Res.* **34**, 267–285 (1987).
19. Wakeham, S. G., Lee, C., Farrington, J. W. & Gagosian, R. B. *Deep Sea Res.* **31**, 509–528 (1984).
20. Amy, P. S., Caldwell, B. A., Soeldner, A. H., Morita, R. Y. & Albright, L. J. *Can. J. fish. aquat. Sci.* **44**, 1135–1142 (1987).
21. Billen, G. in *Microbial Enzymes in Aquatic Environments* (ed. Chróst, R. J.) 123–143 (Springer, New York, 1991).
22. Hoppe, H.-G., Kim, S.-J. & Gocke, K. *Appl. Envir. Microbiol.* **54**, 784–790 (1988).
23. Proctor, L. M. & Fuhrman, J. A. *Nature* **343**, 60–62 (1990).
24. McCave, I. N. *Deep Sea Res.* **31**, 329–352 (1984).
25. Koike, I., Shigemitsu, H., Kazuki, T. & Kogure, K. *Nature* **345**, 242–244 (1990).
26. Longhurst, A. R. *et al. Deep Sea Res.* **39**, 1–7 (1992).
27. Bjørnsen, P. K. *Mar. Ecol. Prog. Ser.* **30**, 191–196 (1986).

ACKNOWLEDGEMENTS. We thank C. Gotschalk, K. Crocker, U. Passow, L. Dilling, D. Campbell and D. Kushner for help in collecting the aggregates, and M. Silver, H. Ducklow and G. Steward for helpful criticism. This work was supported by grants from the US NSF and ONR (F.A. and A.L.A.), and Deutsche Forschungsgemeinschaft (M.S.).

Behavioural evidence for use of a light-dependent magnetoreception mechanism by a vertebrate

John B. Phillips & S. Chris Borland

Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

THE Earth's magnetic field provides an important source of directional information for terrestrial organisms^{1–5}, but the sensory receptor or receptors responsible for magnetic field detection have yet to be identified. Theoretical models of the mechanism of magnetoreception have implicated specialized photoreceptors^{6,7}. The proposed mechanisms would amplify the weak interaction of the geomagnetic field with a single electron spin to the level of photon detection, resulting in a modulation of the photoreceptor response to light. Although behavioural^{8,9} and neurophysiological^{10–12} studies have established a link between magnetic field sensitivity and the visual system, definitive evidence for the use of a light-dependent magnetoreception mechanism has been lacking. Here we show that magnetic compass orientation in a semiaquatic salamander is affected by the wavelength of light¹³, and that this wavelength-dependence is due to a direct effect of light on the underlying magnetoreception mechanism.

Male eastern red-spotted newts (*Notophthalmus viridescens*) were trained to give a consistent magnetic compass response by placing them in water-filled, outdoor tanks with shores at one end¹³. An abrupt increase in water temperature in the training tanks was used to trigger unimodal shoreward magnetic compass orientation¹⁴. Directional responses were observed in an indoor arena in which the horizontal alignment of the magnetic field could be varied¹³. Individual newts were tested in one of four symmetrical alignments of the magnetic field, and the bearings were pooled after being normalized with respect to field alignment¹³. Newts trained under natural (full spectrum) skylight were tested in the indoor arena either under full spectrum light or under one of several discrete wavelength bands of roughly equal quantal content (Fig. 1).

After training under natural light, controls tested under full-spectrum light oriented towards the shore (Fig. 1a, c, e). Newts tested under wavelengths centred at 400 and 450 nm also exhibited shoreward orientation (Fig. 1b), and were indistinguishable from the corresponding controls (Fig. 1a). In contrast, newts tested under 500, 550 and 600 nm light (Fig. 1d) oriented to about 90° anticlockwise of the shoreward direction and differed significantly from full-spectrum controls (Fig. 1c).

The wavelength-dependent shift in orientation shown (Fig. 1a–d) and evidence that newts are randomly oriented with respect to the magnetic field in the absence of visible light¹⁵, do not rule out a nonspecific effect of light on the newts' behaviour. For example, under certain wavelengths of light newts might

be motivated to move parallel to shore and therefore to orient perpendicular to the shoreward direction (as in Fig. 1d). To determine whether light has a direct effect on the magnetic compass, we trained newts in outdoor tanks covered with sheets of yellow gel filter that only transmitted wavelengths above 500 nm. Newts were then tested under either full-spectrum or long-wavelength ($\lambda > 500$ nm) light. Alternative predictions are shown in Fig. 2.

After training under long-wavelength light, newts tested under full-spectrum light oriented roughly 90° clockwise of the shoreward direction (Fig. 1g), whereas newts tested under long-wavelength light oriented towards the shore (Fig. 1h; also Table 1). These data indicate that directional information obtained from the magnetic compass under long-wavelength light is rotated 90° anticlockwise relative to that obtained under full-spectrum light (Fig. 2). The wavelength dependence of magnetic orientation in newts is therefore caused by a direct effect of light on the magnetic compass, rather than by a nonspecific effect on behaviour. Furthermore, the differential effects of short- and long-wavelength light suggest that two spectral mechanisms contribute to this response.

Magnetic compass orientation of terrestrial vertebrates is influenced by the slope or inclination, but not the polarity, of the magnetic field, indicating that the magnetic compass shows a bimodal or 'axial' pattern of response^{5,16,17}. If the inputs that generate such a bimodal pattern are inverted, the pattern will seem to undergo a 90° rotation (Fig. 3). We hypothesize, therefore, that the two spectral mechanisms contributing to the light-dependent magnetic compass produce inverse or 'complementary' patterns of response.

To test this hypothesis (Fig. 3) we observed the orientation of newts under 475-nm light, which lies between the spectral regions producing normal (≤ 450 nm; Fig. 1b) and shifted (≥ 500 nm; Fig. 1d) magnetic orientation, and should therefore be absorbed more or less equally by the two spectral mechanisms. We predicted that under 475-nm light the complementary

TABLE 1 Orientation data

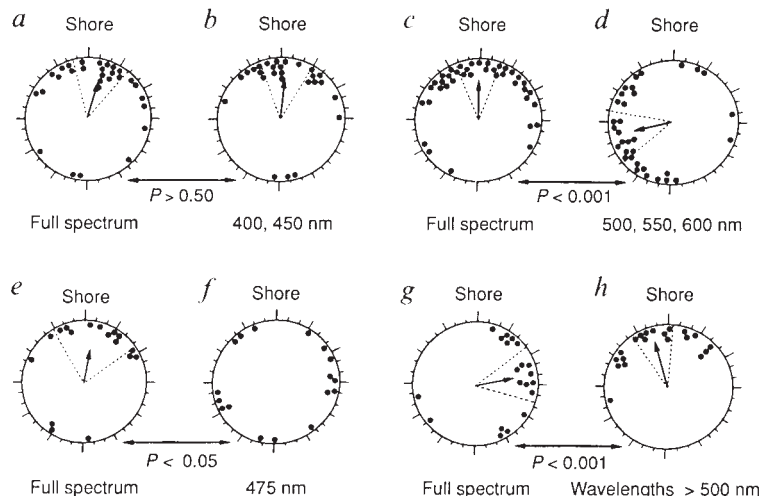
Wavelength/ condition	MVB	r	n	P_R	U^2	P_W
Full spectrum training						
400 nm	2°	0.48	9	n.s.	0.061	n.s.
Full spectrum	344°	0.47	9	n.s.		
400 nm (bimodal)	179–359°	0.72	9	≦0.01	0.102	n.s.
Full spectrum (bimodal)	180–360°	0.61	9	≦0.05		
450 nm*	7°	0.68	18	≦0.001	0.053	n.s.
Full spectrum	22°	0.64	20	≦0.001		
400, 450 nm	5°	0.61	27	≦0.001	0.060	n.s.
Full spectrum	13°	0.56	29	≦0.001		
475 nm	97°	0.07	17	n.s.	0.209	≦0.05
Full spectrum	11°	0.49	17	≦0.05		
500 nm	253°	0.42	14	≦0.1	0.232	≦0.05
Full spectrum	348°	0.64	14	≦0.01		
550 nm	268°	0.49	12	≦0.05	0.211	≦0.05
Full spectrum	14°	0.50	16	≦0.01		
600 nm	241°	0.82	7	≦0.01	0.245	≦0.05
Full spectrum	350°	0.66	9	≦0.01		
500, 550, 600 nm	254°	0.52	33	≦0.001	0.603	≦0.001
Full spectrum	358°	0.58	39	≦0.001		
400, 450 nm	5°	0.61	27	≦0.001	0.694	≦0.001
500, 550, 600 nm	254°	0.52	33	≦0.001		
Long-wavelength training						
Long wavelength	344°	0.76	20	≦0.001	0.521	≦0.001
Full spectrum	79°	0.59	20	≦0.001		

MVB, mean vector bearing of pooled distribution of magnetic bearings rotated so that shore directions are at 0°; *r*, mean vector length; *P_R*, probability, Rayleigh test²⁴; *U*², Watson *U*² statistic; *P_W*, probability, Watson *U*² test²⁴; n.s., not significant ($P \geq 0.1$).

* Data from ref. 13.

FIG. 1 Each pair of distributions show data from newts tested alternately under either full-spectrum light (left) or wavelengths from a specific spectral region (right). Shoreward magnetic orientation of newts, trained under skylight (statistics in Table 1). *a, c, e*, Controls tested under full-spectrum light oriented towards the shore. *b*, Newts tested under 400 and 450 nm light also oriented towards the shore, and were indistinguishable from controls. *d*, Newts tested under 500, 550 and 600 nm light oriented approximately 90° anticlockwise to the shore, and differed significantly from full-spectrum controls. There were no differences in orientation between newts tested under 400 and 450 nm light, between newts tested under 500, 550 and 600 nm light, or between the corresponding controls ($P > 0.20$, Watson U^2). *f*, Newts tested under 475 nm were randomly distributed with respect to the magnetic field and differed significantly from full spectrum controls. Shoreward magnetic orientation of newts trained under long-wavelength light (statistics in Table 1). *g*, Newts tested under full-spectrum light oriented about 90° clockwise of the shore. *h*, Newts tested under wavelengths greater than 500 nm were oriented in the shoreward magnetic direction. The two distributions were significantly different. The newts tested under full-spectrum light after long-wavelength training also differed from newts tested under full-spectrum light after full-spectrum training (compare *a* and *g*: Watson $U^2 = 0.347$, $P < 0.01$; compare *c* and *g*: Watson $U^2 = 0.406$, $P < 0.001$; compare *e* and *g*: Watson $U^2 = 0.258$, $P < 0.02$). Arrows at the centre of distributions indicate significant mean vector bearings; lengths of arrows are proportional to the mean vector lengths, r (radius of circle corresponds to $r = 1$). Dashed lines indicate the 95% confidence interval for the mean vector bearing²⁴.

METHODS. Detailed descriptions of methods have been presented in ref. 13. Groups of newts were trained for at least 5 days in water-filled, outdoor tanks with an artificial shore at one end. Immediately after exposure to a rapid elevation of water temperature, newts were tested individually in a visually symmetrical, terrestrial arena located indoors and illuminated by a 150 W xenon lamp¹³. Directional responses were measured at a distance of 20 cm from the centre of the arena. Individual newts were tested only



once, an equal number in each of four magnetic field alignments (magnetic north at north, east, south or west)¹⁴, and the magnetic bearings were pooled relative to the direction towards shore in the training tank. As a consequence of the use of the four symmetrical magnetic field alignments, the pooled distribution retained only the component of orientation that was a consistent magnetic response¹⁴. Discrete wavelengths of light were produced using broad band interference filters (40 ± 2 nm bandwidth). Quantal flux was 12.65 ± 0.1 log quanta per cm^2 per sec for wavelengths 450–600 nm and 12.35 ± 0.1 quanta per cm^2 per sec at 400 nm. In tests in which newts were trained under long-wavelength light, the outdoor tanks were enclosed in sheets of a yellow gel filter (Lee Filters no. 101) which transmitted less than 1.0% of light at wavelengths below 490 nm and less than 0.1% at wavelengths below 470 nm. For testing under long-wavelength light, a filter made of the same yellow gel material was placed in the light path to the arena.

patterns of response produced by the two spectral mechanisms should cancel each other out, and eliminate the newt's ability to detect the magnetic field. Figure 1*f* shows that the newts' magnetic compass response was indeed completely abolished.

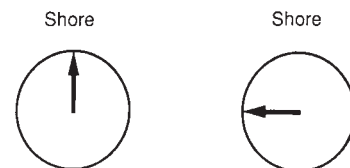
We can only speculate about the nature of the spectral mechanisms. One possibility is that there are two types of light-dependent magnetoreceptors that differ in spectral sensitivity and provide antagonistic inputs to a second-order cell¹³ (similar to the spectral opponent units found in amphibians^{18–20}). An alternative is that there are two spectral mechanisms present in the same cell which have antagonistic effects on the cell's response, like those that seem to occur in

some photoreceptors in the pineal complex of amphibians²¹ and lizards (G. Engbreton, personal communication). This alternative is particularly interesting as the pineal gland has been implicated as an independent site of magnetic field transduction in birds²².

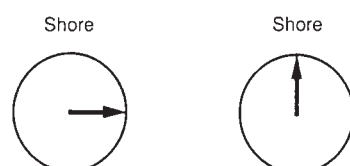
A potential problem with the mechanism demonstrated here is that most light encountered under natural conditions is broad-band and therefore should excite both spectral mechanisms. As under 475-nm light (Fig. 1*f*), 'balanced' activation of the two spectral mechanisms by broad-band light should eliminate, or at least seriously degrade, the directional information obtained from the magnetic compass.

FIG. 2 Predicted magnetic orientation after training under long-wavelength light. Alternatively 1. If the wavelength dependence of shoreward magnetic orientation is due to a nonspecific effect on behaviour, newts tested under long-wavelength and full spectrum light should exhibit the same responses shown by newts trained under full spectrum light, that is, newts tested under full spectrum light should orient toward shore (Fig. 1*a, c, e*), while newts tested under long-wavelength light should orient 90° anticlockwise of shore (Fig. 1*d*). Alternatively 2. If long-wavelength light causes a 90° anticlockwise rotation of directional information obtained from the magnetic compass, newts trained under long-wavelength light should learn the direction of shore with respect to the rotated directional information. When subsequently tested under full-spectrum light, the directional information obtained from the magnetic compass would be rotated in a clockwise direction from that experienced during training. As a consequence, they should orient 90° clockwise of the shoreward direction. In contrast, when tested under long-wavelength light, newts will be exposed to the same magnetic compass information experienced during training and should therefore orient in the correct shoreward direction.

Alternative 1: Nonspecific effect on behaviour
Testing: Full spectrum Long wavelength



Alternative 2: Direct effect on magnetic compass
Testing: Full spectrum Long wavelength



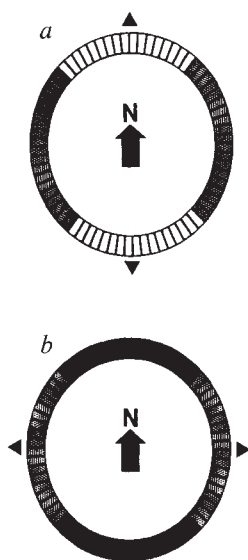


FIG. 3 A hypothetical magnetoreception system consisting of a circular array of receptors (small rectangles). *a*, Under short-wavelength light, receptors aligned within a certain angle of either end of the magnetic axis exhibit an increase in response (white rectangles) relative to receptors in alignments that are not affected by the magnetic field (hatched rectangles). Arrows at the edge of the array indicate the axis with the highest level of response. *b*, Under long-wavelength light, receptors in alignments that are affected by the magnetic field exhibit a decrease in response (dark region) relative to receptors in alignments that are unaffected by the magnetic field (hatched rectangles). The axes with the highest level of response differ by 90° in *a* and *b*. This hypothesis does not predict the direction along the magnetic axis that will be interpreted as magnetic north. This direction will depend on the particular set of 'rules' that an organism uses to derive directional information from the magnetic field. Shoreward-orienting newts⁵ and nocturnally migrating birds^{16,17} use dip angle or inclination to distinguish between the two ends of the magnetic field axis, but different sets of rules may be used by other organisms. Indeed, recent experiments carried out in our laboratory demonstrate that magnetic compass orientation in *Drosophila melanogaster* undergoes a 90° clockwise rotation under long-wavelength light (J.B.P. and O. Sayeed, manuscript in preparation), in contrast to the 90° anticlockwise rotation exhibited by newts (Fig. 1).

The strong magnetic orientation exhibited by newts under full-spectrum light (Fig. 1*a, c, e*) indicates that broad-band light does not produce the expected degradation of response. Balanced activation seems to be avoided by a reduction in the sensitivity of the long-wavelength spectral mechanism, because the response of newts tested under full-spectrum light is indistinguishable from that of newts tested under short wavelength (400 and 450 nm) light (Fig. 1*b*; Table 1).

Reduction in the sensitivity of the long-wavelength spectral mechanism could result from the selective absorption of long-wavelength light by a screening pigment, as occurs in the retinula 7y cells of the flies *Musca* and *Calliphora* which contain two spectral mechanisms within the same cell—a sensitizing pigment absorbing at short wavelengths and a photopigment absorbing at long wavelengths²³. Alternatively, the reduction in long-wavelength sensitivity may result from a matched spectral input that inhibits the long-wavelength spectral mechanism. In an amphibian, units have been identified in the retina that receive an excitatory input from a short-wavelength spectral mechanism, and matching excitatory and inhibitory inputs from two long-wavelength mechanisms with similar spectral sensitivities. The antagonistic long-wavelength inputs cancel each other out, except for transient on and off responses, resulting in a substantial reduction in the response to long wavelength light¹⁹.

Further work is needed to identify and characterize the physical and physiological processes responsible for the wavelength-

dependence of the newt's magnetic compass. A better understanding of this sensory system is likely to give new insight into the competing selective pressures and functional constraints that have shaped the vertebrate visual system. □

Received 13 April; accepted 21 July 1992.

1. Wiltischko, W. & Wiltischko, R. *Experientia* **46**, 342–352 (1990).
2. Burda, H., Mayrhoth, S., Westenberger, T., Wiltischko, R. & Wiltischko, W. *Experientia* **46**, 528–530 (1990).
3. Sinsch, U. *J. comp. Physiol.* **161**, 715–727 (1987).
4. Quinn, T. P. & Brannon, E. L. *J. comp. Physiol.* **147**, 547–552 (1982).
5. Phillips, J. B. *Science* **233**, 765–767 (1986).
6. Leask, M. J. M. *Nature* **267**, 144–145 (1977).
7. Schulten, K. *Adv. Solid State Phys.* **22**, 61–83 (1982).
8. Wiltischko, W. & Wiltischko, R. *Nature* **291**, 433–434 (1981).
9. Olcese, J., Reuss, S. & Semm, P. *Life Sci.* **42**, 605–613 (1988).
10. Semm, P., Nohr, D., Demaine, C. & Wiltischko, W. *J. comp. Physiol.* **155**, 283–288 (1984).
11. Semm, P. & Demaine, C. *J. comp. Physiol.* **159**, 619–625 (1986).
12. Phillips, J. B. & Borland, S. C. *Ethol. Ecol. Evol.* **4**, 33–42 (1992).
13. Phillips, J. B. *J. comp. Physiol.* **158**, 103–109 (1985).
14. Phillips, J. B. & Borland, S. C. *Anim. Behav.* (in the press).
15. Wiltischko, W. & Wiltischko, R. *Science* **176**, 62–64 (1972).
16. Beason, R. C. *Ethology* **81**, 291–299 (1989).
17. Reuter, R. & Virtanen, K. *J. comp. Physiol.* **109**, 337–343 (1976).
18. Backstrom, A. C. & Reuter, T. *J. Physiol. Lond.* **246**, 79–107 (1975).
19. Tempel, P., Himstedt, W. & Steinke, S. *Zool. Jb. Zool.* **86**, 401–412 (1982).
20. Dodt, E. & Heerd, E. *J. Neurophysiol.* **25**, 405–429 (1963).
21. Demaine, C. & Semm, P. *Neurosci. Lett.* **62**, 119–121 (1985).
22. Kirschfeld, K., Hardie, R., Lentz, G. & Vogt, K. *J. comp. Physiol.* **162**, 421–433.
23. Batschelet, E. *Circular Statistics in Biology* (Academic, New York, 1981).

ATP receptor-mediated synaptic currents in the central nervous system

Frances A. Edwards*, Alasdair J. Gibb & David Colquhoun

Department of Pharmacology, University College London, Gower Street, London WC1E 6BT, UK

UNTIL now, the only well documented, fast excitatory neurotransmitter in the brain has been glutamate. Although there is evidence for adenosine 5'-triphosphate (ATP) acting as a transmitter in the peripheral nervous system^{1–4}, suggestions for such a role in the central nervous system^{1,5–11} have so far not been supported by any direct evidence. Here we report the recording of evoked and miniature synaptic currents in the rat medial habenula. The fast rise time of the currents showed that they were mediated by a ligand-activated ion channel rather than a second messenger system, thus limiting the known transmitter candidates. Evidence was found for the presence on the cells of glutamate, γ -aminobutyric acid, acetylcholine and ATP receptors, but not for 5-hydroxytryptamine (5HT₃) or glycine receptors. The evoked currents were unaffected by blockers of glutamate, γ -aminobutyric acid or acetylcholine receptors but were blocked by the ATP receptor-blocker, suramin and the desensitizing ATP receptor-agonist α, β -methylene-ATP. Our evidence identifies for the first time synaptic currents in the brain, mediated directly by ATP receptors.

Whole-cell patch-clamp recordings from medial habenula slices (Fig. 1*a*) show frequent spontaneous synaptic currents. This activity (in the presence or absence of tetrodotoxin (TTX)) is greatly decreased if the glutamate receptor-blockers, 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX, 5 μ M, non-*N*-methyl-D-aspartate (NMDA) receptors) and 7-chlorokynurenatate (4 μ M, NMDA receptors) or the γ -aminobutyric acid (GABA_A) receptor-blocker, bicuculline (5 μ M) are added to the bath solution. Thus, like most other central neurons, these cells receive synaptic inputs mediated by glutamate and GABA. Initially we investi-

* Present address: Department of Pharmacology, University of Sydney, New South Wales 2006, Australia.

gated whether these cells also had nicotinic synapses and thus for all further experiments glutamate and GABA receptor-blockers and the muscarinic blocker, atropine ($1\text{ }\mu\text{M}$), were included in the bath solution, generally being added at least 30 min before recording.

In this blocking solution, it was still possible to evoke inward synaptic currents when stimulation pulses were applied in the nearby tissue (Fig. 1*b*). Occurrence of evoked currents was prevented by addition of the sodium channel blocker, TTX ($1\text{ }\mu\text{M}$), to block action potentials, showing that they were due to stimulation of a presynaptic fibre. The latency was usually about 2 ms, although in a few cells the response consisted of a mixture both of currents with short latency and of currents with somewhat longer latencies (3–4 ms), suggesting that in most cases the evoked currents were the result of stimulation of a monosynaptic pathway. The rise time of evoked synaptic currents varied in different cells and sometimes within a cell, presumably reflecting the position of the stimulated synapses on the cell. They were, however, generally less than 2 ms and

sometimes well under 1 ms. The decay phase of averaged currents, aligned by the stimulus artefact could be fitted well with a single exponential with a time constant of $17.4 \pm 3.0\text{ ms}$ (mean $\pm$ s.d., 6 cells).

The fast rise time of the evoked currents suggested that they were mediated by an ion channel directly linked to a receptor, rather than a second messenger system. With glutamate and GABA_A receptors blocked, the only remaining ligand-gated channels so far identified that could mediate the evoked currents are acetylcholine (nicotinic), ATP, 5-hydroxytryptamine (5HT₃)

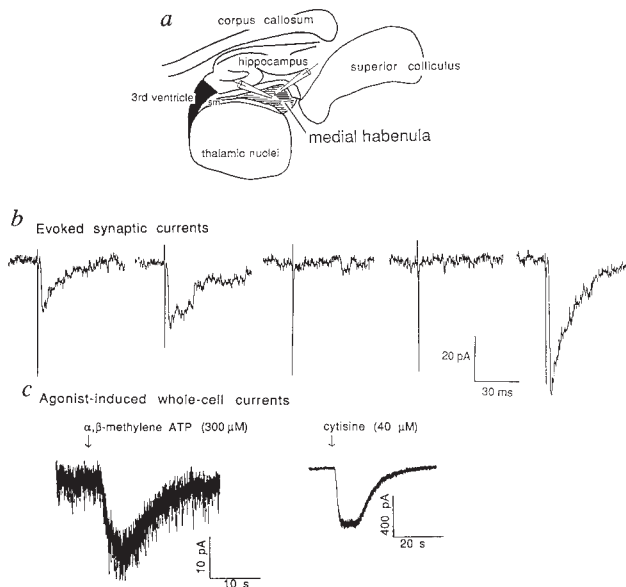


FIG. 1 Currents in the medial habenula. *a*, The position of the medial habenula with respect to other brain regions. During preparation, a full parasagittal brain slice ($200\text{ }\mu\text{m}$) was cut, the medial habenula was then dissected free retaining only the surrounding stria medullaris (sm). Position of recording and stimulating electrodes show a typical arrangement, though no particular part of the medial habenula was preferred. *b*, Responses to consecutive stimuli at -70 mV in normal Krebs solution containing blockers of glutamate, muscarinic and GABA receptors. All synaptic currents in this and the following figure were digitized at 10 kHz and filtered at 3 kHz (-3 dB , 8 pole Bessel filter). *c*, Whole-cell responses due to local application of ATP and nicotinic receptor-agonists.

METHODS. The technique for making patch-clamp recordings in brain slices was as previously described²⁴. Incubation ($35\text{ }^{\circ}\text{C}$, $>30\text{ min}$) and recording ($21\text{--}25\text{ }^{\circ}\text{C}$) were in normal Krebs solution (containing in mM: NaCl 125, KCl 2.5, NaHCO_3 26, NaH_2PO_4 1.25, glucose 25, CaCl_2 2, MgCl_2 1 bubbled with 5%/95% CO_2/O_2). Intracellular solution contained (in mM): KCl 140; EGTA 10; HEPES 10; CaCl_2 1; Mg-ATP 2. Electrodes ($3\text{--}6\text{ M}\Omega$) were pulled from either borosilicate or aluminosilicate glass. Trimethaphan was provided by M. Farrant. Synaptic currents were evoked by delivering short, high-voltage pulses ($5\text{--}85\text{ V}$, $200\text{ }\mu\text{s}$) through a stimulation pipette ($1\text{ }\mu\text{m}$ tip diameter) which was moved around the tissue until a response was seen in the recorded cell (*b*). Data were recorded through an EPC-7 patch-clamp amplifier (List, Darmstadt) on FM tape (Racal 4, DC 5 kHz) and digitized with a CED 1401-plus interface (Cambridge Electronic Design).

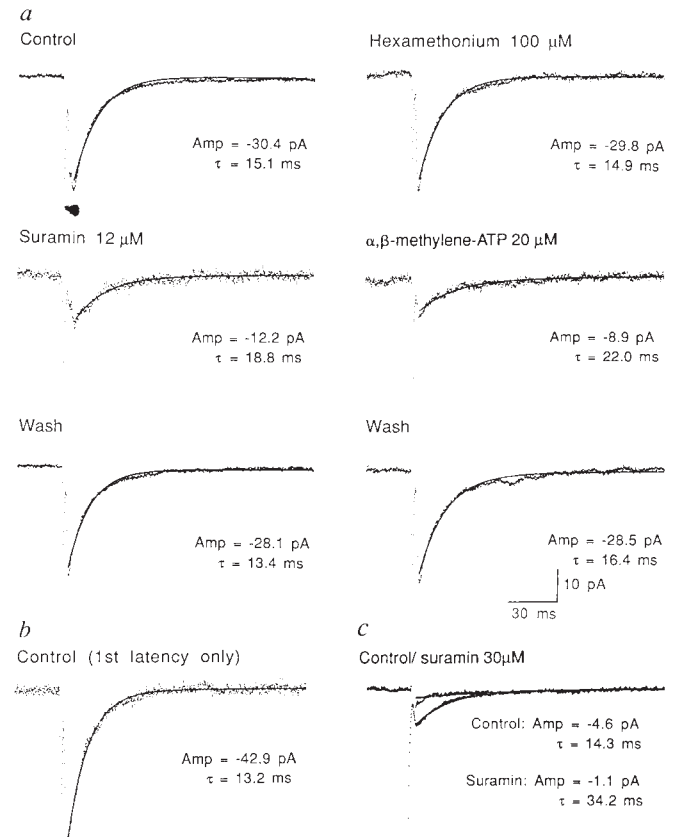
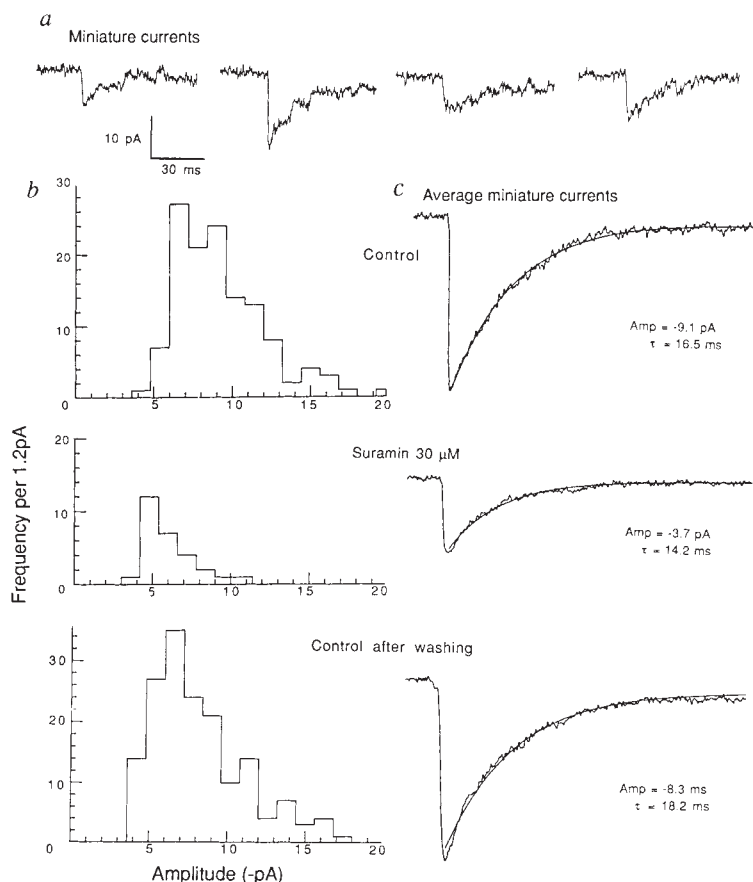


FIG. 2 Effects of suramin, α,β -methylene ATP and hexamethonium on evoked synaptic currents. *a*, Averages of consecutive responses from the same cell, reading from top to bottom of the left and then the right column (-70 mV , $n=100\text{--}360$ responses, including apparent failures). The blocking effect of suramin ($12\text{ }\mu\text{M}$) was completely reversible. Hexamethonium ($100\text{ }\mu\text{M}$), included in the washing solution, had no effect. After washing, addition of α,β -methylene ATP ($20\text{ }\mu\text{M}$) partially blocked the evoked currents. This effect was also reversible. In the same cell (not shown) TTX blocked synaptic responses reversibly and hexamethonium in combination with suramin blocked responses to the same degree as suramin alone. Mixed latencies in this cell resulted in a poor single exponential fit of the decay phase. *b*, Average of 30 synaptic currents in the same cell as above, selected for shortest latency group ($\sim 2\text{ ms}$). Selection resulted in a much improved fit with a single exponential function. Apparent failures were not included in this average. Whether the first or second latency group was selected did not affect the percentage block by suramin. *c*, Superimposed averages of responses in control solution and in the presence of suramin ($30\text{ }\mu\text{M}$). The relatively small amplitude of the control average in this cell is due to a higher apparent failure rate than in the cell above. In individual traces, response to stimulation was never detectable in the presence of suramin ($30\text{ }\mu\text{M}$). Note that the residual current left in the presence of suramin could not be fitted by a comparable exponential to control currents and may be related to the stimulus artefact. Current remaining in TTX was similar. In one cell in which such a block was seen by $30\text{ }\mu\text{M}$ suramin. Suramin ($3\text{ }\mu\text{M}$) reversibly increased the amplitude of evoked synaptic currents. This effect was not pursued in the present study. The scale bar refers to all sections of the figure. 'Amp' and ' τ ' refer to the amplitude and decay time constant of the single exponential fits shown as solid lines.

FIG. 3 Effect of suramin (30 μM) on miniature synaptic currents. *a*, Typical miniature synaptic currents (-70 mV in Krebs solution containing TTX, CNQX, 7-Cl-kynurene, bicuculline and atropine). Relatively high concentrations of both K^+ (10 mM) and Ca^{2+} (5 mM) were added to the solution to increase frequency. *b*, Amplitude histograms of miniature synaptic currents. Top, amplitudes of currents recorded over an 8 min period in control solution (as above). Middle, suramin (30 μM) was added to the bath and currents were recorded for a period of 6 min 40 s. Lower, amplitudes of miniature currents recorded for a period of 8 min after 5 min of washing with control solution. Events, sampled at 10 kHz, filtered at 1 kHz (-3 dB) were accepted for analysis if they crossed a threshold of 3–4 times the baseline r.m.s. noise level (0.77–0.91 pA). Note that this criterion may allow acceptance near the threshold level (for example in the range 3.5–5 pA) of some false events and will exclude any real events of less than 3.5 pA. (Fig. 3*b, c*). Before addition of suramin, the frequency of detected miniatures was 21 ± 3 per min with an average amplitude of 8.6 ± 1.1 pA. The distribution was skewed. In the presence of suramin the frequency of detectable currents was greatly reduced (8.5 ± 1 per min, average amplitude 5.8 ± 0.3 pA) with most of the remaining events barely crossing the detection threshold. After washout of suramin with control solution, the frequency and amplitude of detected events returned towards control levels (14.7 ± 3 per min, average amplitude 7.3 ± 0.9 pA). *c*, Averages of consecutive miniature synaptic currents in a different cell. Upper, average of 24 currents in control solution; middle, average of 90 miniature currents in the presence of suramin (30 μM); lower, average of 30 currents after washing out suramin for 10 min with the above control solution. Currents were lined up for averaging using the steepest point of their rising phase.



and glycine receptor/channels. We used bath or local application of agonists for these receptors to investigate whether the receptors were present on the cells. Bath application of glycine (100 μM) resulted in no measurable whole-cell current. Bath or local application of 5HT (10 μM) resulted in a small (~ 20 pA), delayed, outward current showing that the agonist reached the cell and presumably stimulated a second messenger-linked 5HT₁ or 5HT₂ receptor. Thus there was no evidence for the presence of glycine or 5HT₃ receptors on these cells. Evoked synaptic currents were not affected in the presence of these agonists, showing that if the lack of whole-cell response was due to desensitization, these were nevertheless not the receptors mediating the evoked synaptic currents.

Acetylcholine (ACh) and ATP are harder to apply exogenously because both agonists are broken down very rapidly in the slice. When ACh (100 μM) was applied in the bath solution or locally, a whole-cell current was seen only if the slice was previously incubated with an anticholinesterase, methane sulphonyl fluoride (100 μM , 15 min). The presence of nicotinic receptor/channels was further confirmed by local application of the nicotinic agonist, cytisine (40 μM , Fig. 1*c*). We attempted to apply ATP at high concentration (100 μM) in the bath solution to evoke a whole-cell current. Unfortunately there is as yet no blocker for the ecto-enzymes which hydrolyse ATP¹² and no whole-cell response was seen. We thus used local application of the non-hydrolysable ATP analogue α, β -methylene ATP (300 μM). Small inward currents were evoked (10–100 pA, Fig. 1*c*) in all cells tested ($n = 10$). Thus in addition to glutamate and GABA-gated ion channels, medial habenula neurons contain nicotinic and ATP receptors. Antagonists were applied to distinguish which of these two receptor/channel types was likely to underlie the evoked currents seen in the presence of glutamate and GABA receptor-blockers.

Various nicotinic blockers were tested (hexamethonium 100 μM (Fig. 2*a*), mecamylamine 0.1, 1.0 μM and trimethaphan

5 μM) that had no effect on amplitude or time course of the evoked currents. In contrast, addition of the ATP-receptor blocker suramin to the bath solution partially (12 μM , 6 cells, Fig. 2*a*) or almost completely (30 μM , 4 cells, Fig. 2*c*) blocked the response. These effects were reversible (Fig. 2*a*). As previously shown in coeliac ganglia^{3,4} and PC12 cells¹³, suramin (30 μM) had no measurable effect on nicotinic receptor-mediated whole-cell currents. Evoked currents were also reversibly blocked by addition to the bath of the desensitizing ATP receptor-agonist, α, β -methylene ATP (20 μM , 3 cells, Fig. 2*a*). Note that stimulation of the incoming nerve bundle, stria medullaris (sm, Fig. 1*a*) resulted in synaptic currents which were sensitive to CNQX, but insensitive to any of the other antagonists tested, including suramin. These currents were not further investigated.

Miniature synaptic currents are thought to occur as a result of spontaneous random release of transmitter packets from presynaptic terminals. To test whether suramin also blocked miniature currents, spontaneously occurring currents were recorded in the presence of TTX (1 μM , Fig. 3*a*). Suramin (30 μM , 3 cells) reduced both the frequency of occurrence and the mean amplitude of miniature synaptic currents (Fig. 3*b, c*). The suramin-sensitive miniatures showed a very similar time course to the evoked currents (Fig. 3*c* versus Fig. 2*b*).

To confirm the presence of ATP receptors on the cells we looked for ATP-mediated single-channel currents on outside-out patches pulled from medial habenula neurons. In contrast to the consistent whole-cell response to applied agonist, application of ATP or α, β -methylene ATP to six outside-out patches, evoked single-channel currents in only one recording. Although this confirms the presence of ATP receptor/channels on these cells, it suggests that the overall density is low, with receptors possibly being concentrated at the synapse. This is also suggested by the small whole-cell currents seen in response to local application of exogenous agonist, but this apparently small response could also be explained by desensitization of a large proportion

of receptors. As currents were only recorded in one patch, analysis of channel properties was not included in this study but features such as their apparent conductance (around 20 pS) have at least some similarity to the P_{2x} receptor identified in peripheral tissues^{2,14}.

Previous evidence for ATP receptor-mediated effects in the brain include modulation of general synaptic activity^{2,5-11} and Ca^{2+} -dependent release of ATP from synaptosomal fragments of rat brain¹⁵⁻¹⁸. ATP-mediated single-channel currents have been recorded in rat sensory neurons¹⁹, PC-12 cells²⁰ and cardiac ganglion cells²¹. Note that in an electron microscope study of the medial habenula, Tokunaga and Otani (1978)²² observed four types of terminal. Types 1 and 2 they proposed to be classical excitatory and inhibitory synapses, respectively, and could correspond to the glutamatergic and GABAergic synaptic currents recorded in these cells. Type 3 they described as "endings with large dense core vesicles [which] seem to contain a neurotransmitter which remains to be specified cytochemically." They continue "Burnstock described similar endings in peripheral organs and correlated them with purinergic transmission". But despite diverse indirect evidence from previous studies, this is the first time that synaptic currents, carried by ATP receptor/channels have been demonstrated in the central nervous system.

We have demonstrated a novel type of synaptic current in the medial habenula which is carried by a receptor-coupled ion channel. By a combination of eliminating all other known receptor-gated channels and specifically blocking both evoked and miniature currents with the ATP receptor-antagonist, suramin, or the desensitizing agonist, α,β -methylene ATP, we have demonstrated that the synaptic response is due to activation of ATP receptors on the postsynaptic membrane of these cells. In addition, by locally applying α,β -methylene ATP fast enough to avoid complete desensitization, we have been able to demonstrate that local exogenous application of agonist also results in inward currents both in the whole-cell configuration and on an outside-out patch. Unless an as yet unknown transmitter is discovered, that opens receptor-gated ion channels blocked by suramin and desensitized by α,β -methylene ATP, we conclude that ATP is a fast neurotransmitter in the rat medial habenula. The functional significance of this new central transmitter is so far unknown, but its existence in the brain suggests a possible new target for central therapeutic agents²³. □

Isoprenylation in regulation of signal transduction by G-protein-coupled receptor kinases

James Inglese*, Walter J. Koch, Marc G. Caron† & Robert J. Lefkowitz*‡§

* Howard Hughes Medical Institute, Departments of Biochemistry, ‡ Medicine and † Cell Biology, Duke University Medical Centre, Durham, North Carolina 27710, USA

RHODOPSIN kinase¹ and β -adrenergic receptor kinase (β ARK)² are related members of a serine/threonine kinase family that specifically initiate deactivation of G-protein-coupled receptors. After stimulus-mediated receptor activation, these cytoplasmic kinases translocate to the plasma membrane^{3,4}. Here we show that the molecular basis for this event involves a class of unsaturated lipids called isoprenoids. Covalent modification *in vivo* of rhodopsin kinase by a 15-C (farnesyl) isoprenoid⁵ enables the kinase to anchor to photon-activated rhodopsin. Mutations that alter or eliminate the isoprenoid, fully disable light-specific Rhodopsin kinase translocation. Other receptor kinases (such as β ARK), which lack an intrinsic lipid, are activated⁶ on exposure to brain $\beta\gamma$ subunits of the signal-transducing G proteins, the γ subunit of which bears a 20-C (geranylgeranyl) isoprenoid^{7,8}. Using chimaeric β ARKs that undergo isoprenylation *in vitro*, we demonstrate that membrane association and activation of these kinases can occur in the absence of $\beta\gamma$. These results indicate that rhodopsin kinase (by means of an integral isoprenoid) and β ARK (through its association with $\beta\gamma$) both rely on the function of isoprenyl moieties for their translocation and activity, illustrating distinct, though related, modes of biological regulation of receptor function.

Protein isoprenylation is a post-translational modification in which a farnesyl or geranylgeranyl isoprenoid, derived from the mevalonate pathway, is attached to a carboxy-terminal cysteine residue^{9,10}. The biological significance of this modification is not fully understood, but in the best known case, that of the *ras* oncogene product (*ras*), the farnesyl moiety is required for transforming activity and membrane association¹¹⁻¹⁴. Because the role of *ras* in cellular signalling is unclear¹⁵, however, the precise function of isoprenylation in this system has been difficult to establish. The dependence of rhodopsin kinase activity on farnesylation⁵, and the activation of a muscarinic receptor kinase⁶ and β ARK¹⁶ by $\beta\gamma$, led us to investigate whether isoprenylation might have a general role in regulating the activity of these kinases and, if so, what the regulation mechanism might be. We constructed three forms each of rhodopsin kinase and β ARK; one bearing a farnesyl and one a geranylgeranyl isoprenoid and one non-isoprenylated (Fig. 1).

Farnesylated and geranylgeranylated rhodopsin kinase have about equal ability to phosphorylate light-bleached rhodopsin in rod outer-segment (ROS) membranes, whereas the non-isoprenylated form has significantly diminished activity⁵ (Fig. 2a). Virtually identical results were obtained when the β_2 -adrenergic receptor, reconstituted in vesicles, was used as the substrate (data not shown). We hypothesized that the molecular basis for these differences might become apparent by examination of the stimulus-mediated movement of receptor kinase from cytosol to membrane, the process of translocation. Rhodopsin kinase is restricted to the retina¹, where the process is initiated by a photon³. β ARK is distributed throughout the brain and peripheral tissues², where β -adrenergic receptor agonists, as well as agents acting on other G-protein-coupled receptors (such as prostaglandin E_1 and somatostatin)^{4,17}, have been shown to initiate translocation, indicating that a greater diversity

Received 14 May; accepted 4 August 1992.

1. Hoyle, C. H. V. & Burnstock, G. in *Adenosine in the Nervous System* (ed. Stone, T. W.) Chapter 3 (Academic, London, 1991).
2. Bean, B. *Trends Pharmac. Sci.* **13**, 87-90 (1992).
3. Silinsky, E. M., Gerzanich, V. & Vannier, S. M. *Br. J. Pharmac.* **106**, 762-763 (1992).
4. Evans, R. J., Derkach, V. & Surprenant, A. *Nature* **357**, 503-505 (1992).
5. Phillips, J. W., Kostopoulos, G. K. & Limacher, J. J. *Can. J. Physiol.* **52**, 1226-1229 (1974).
6. Fyfe, R. E. W. & Perl, E. R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6890-6893 (1984).
7. Salter, M. W. & Henry, J. L. *Neuroscience* **15**, 815-825 (1985).
8. Jahr, C. E. & Jessel, T. M. *Nature* **304**, 730-733 (1983).
9. Tschöpel, M., Harms, L., Nörenberg, W. & Illies, P. *Eur. J. Pharm.* **213**, 71-77 (1992).
10. Wierasko, A. & Seyfried, T. N. *Brain Res.* **491**, 356-359 (1989).
11. Inoue, K., Nakazawa, K., Fujimori, K., Wotano, T. & Takanaka, A. *Neurosci. Lett.* **134**, 215-218 (1992).
12. Schubert, P. & Kreutzberg, G. W. *Brain Res.* **168**, 419-424 (1979).
13. Nakazawa, K., Fujimori, K., Takanaka, A. & Inoue, K. *J. Physiol.* **434**, 647-660 (1991).
14. Burnstock, G. & Kennedy, C. *Gen. Pharm.* **16**, 433-440 (1985).
15. White, T. D. *J. Neurochem.* **30**, 329-336 (1976).
16. Potter, P. & White, T. D. *Neuroscience* **5**, 1351-1356 (1980).
17. Richardson, P. J. & Brown, S. J. *Biochem. Soc. Trans.* **14**, 1250-1251 (1986).
18. Wierasko, A., Goldsmith, G. & Seyfried, T. N. *Brain Res.* **485**, 244-250 (1989).
19. Krishtal, O. A., Marchenko, S. M. & Obukhov, A. G. *Neuroscience* **27**, 995-1000 (1988).
20. Neuhäus, R., Reber, B. F. X. & Reuter, H. *J. Neurosci.* **11**, 3984-3990 (1991).
21. Fieber, L. A. & Adams, D. J. *J. Physiol.* **434**, 239-256 (1991).
22. Tokunaga, A. & Otani, K. *Brain Res.* **150**, 600-606 (1978).
23. Williams, M. in *Psychopharmacology: The Third Generation of Progress* Chapter 30 (ed. Meltzer, H. Y.) (Raven Press, New York, 1987).
24. Edwards, F. A., Konnerth, A., Sakmann, B. & Takahashi, T. *Pflügers Arch.* **414**, 600-612 (1989).

ACKNOWLEDGEMENTS. We thank E. Ensor for technical assistance, G. Burnstock for contribution of drugs and invaluable discussion and G. Johnston, M. Christie, J. Black and J. Martin for comments on the manuscript. F.A.E. was supported by a Wellcome Research Fellowship and the laboratory is funded by the British Medical Research Council.

§ To whom correspondence should be addressed.

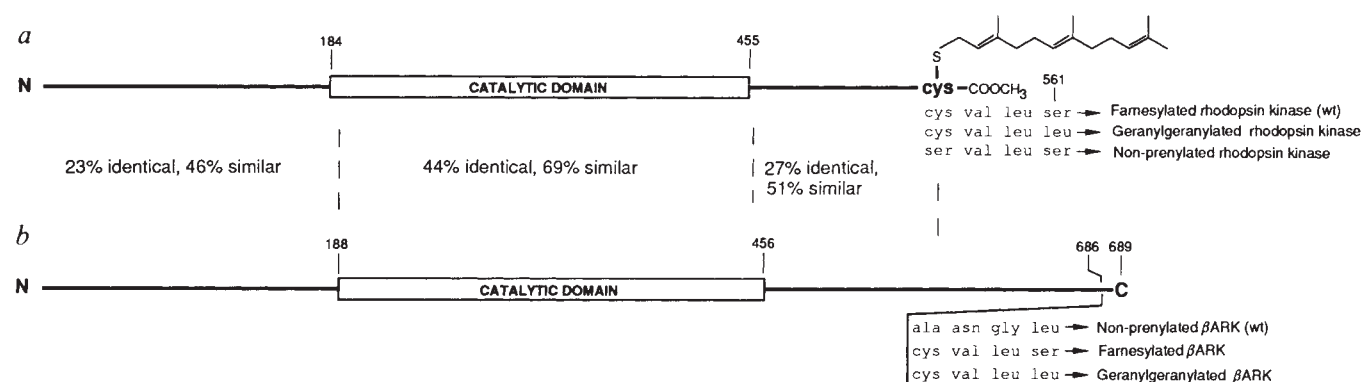
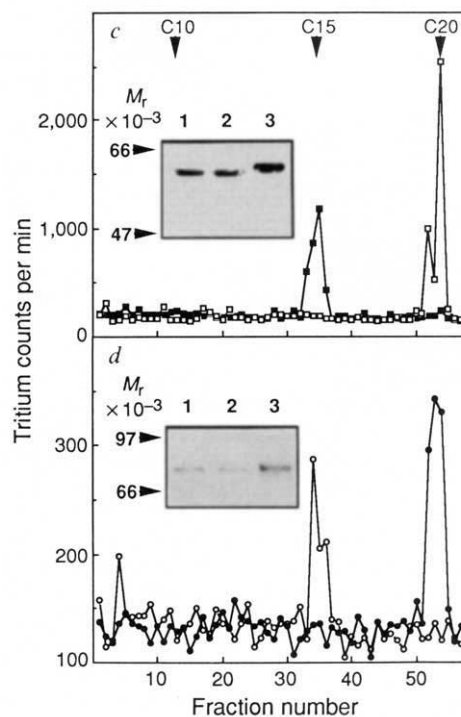


FIG. 1 The G-protein-coupled receptor kinases. *a*, Schematic diagram of rhodopsin kinase and *b*, β ARK illustrating the C terminus of the processed form of the wild-type (wt) enzymes and the sequences of the CAAX motifs (C, cysteine; A, an aliphatic amino acid; and X, the C-terminal amino acid of the protein) of the preprocessed form of the wt enzymes, mutant rhodopsin kinases and chimaeric β ARKs. Isoprenylation specificity is directed by the C-terminal amino acid of the CAAX motif (for example, X=serine directs farnesylation and X=leucine directs geranylgeranylation) through specific isoprenyl transferases^{21–23}. *c*, High-performance liquid chromatography analysis confirming the isoprenoid type modifying farnesylated rhodopsin kinase (■) and geranylgeranylated rhodopsin kinase (□). Arrows at the top indicate elution time of geraniol (C10), farnesol (C15), and geranylgeraniol (C20) standards. *Inset*, Immunoblot of rhodopsin kinase (lane 1) and geranylgeranylated rhodopsin kinase (lane 2) where processed forms migrate at a lower apparent M_r than the non-processed form, non-isoprenylated rhodopsin kinase (lane 3) as previously described⁵. Molecular standards are indicated on the left. *d*, As *c*, except analysis of farnesylated β ARK (○) and geranylgeranylated β ARK (●) derived isoprenyl alcohols. *Inset*, Immunoblot of farnesylated β ARK (lane 1), geranylgeranylated β ARK (lane 2) and wild-type β ARK (lane 3). Confirmation that farnesylated and geranylgeranylated β ARK undergo stoichiometric isoprenylation was obtained by documenting similar specific activities ([³H]mevalonate-derived [³H] incorporation per pmol kinase) to that obtained for farnesylated and geranylgeranylated rhodopsin kinase, respectively. To ensure complete isoprenylation of the kinases, 4 mM mevalonate was added to the growth media. **METHODS.** Bovine rhodopsin kinase and β ARK (ref. 2) complementary DNAs were subcloned into the expression plasmid pCMV5 (ref. 24) and all subsequent mutagenesis used these as templates. Geranylgeranylated rhodopsin kinase, farnesylated β ARK and geranylgeranylated β ARK were constructed using standard polymerase chain reaction (PCR) techniques as previously described for non-isoprenylated rhodopsin kinase (C558S) (ref. 5). PCR primers corresponded to the new CAAX motifs and the amplified mutant cassettes were spliced into the parent plasmids after removal of the wild-type cassettes. COS-7 cells were transfected with the plasmid constructs described above by the diethylaminoethyl Dextran method as described⁵. Cells were grown on 150 cm² dishes in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum and 4 mM mevalonic acid lactone to ensure 100% isoprenylation. Cells were grown for 42 h and then for 6 h in methionine-free DMEM plus 3% fetal calf serum supplemented with 25 μ Ci ml⁻¹ [³⁵S]-methionine. Cells were collected and homogenized in cold lysis buffer (75 mM Tris-HCl pH 7.5, 1 mM magnesium acetate, 1 mM dithiothreitol (DTT), 20% adonitol, 10 μ g ml⁻¹ leupeptin,



20 μ g ml⁻¹ aprotinin and 3 mM phenylmethylsulphonyl fluoride (PMSF). Nuclei were pelleted at 500g and supernatants centrifuged at 350,000g for 20 min at 4 °C. Supernatants were the source of kinases used in all subsequent assays. Isoprenyl group analysis for each kinase containing an intact CAAX box was as described^{5,25}. The expressed kinases, metabolically labelled with [³H]mevalonic acid lactone (NEN), were immunoprecipitated (using anti rhodopsin kinase⁵ or anti- β ARK antibodies²⁶, proteolysed with trypsin, S-methylated with methyl iodide and the resultant sulphonium iodide hydrolysed to release the isoprenyl groups as their corresponding alcohols. HPLC analysis as described^{5,25}. All proteins examined yielded the expected isoprenyl alcohols.

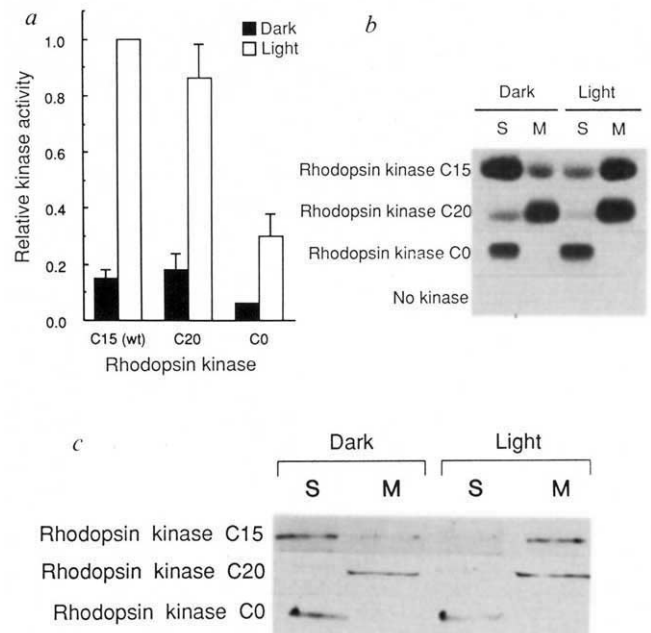
of putative receptor substrates exists for β ARK than for rhodopsin kinase *in vivo*.

We developed an assay to study G-protein-coupled receptor kinase translocation, exploiting the property that wild-type farnesylated rhodopsin kinase will translocate from cytosol to ROS membranes on light-activation. Farnesylated rhodopsin kinase undergoes light-dependent redistribution (Fig. 2b) as indicated by the dramatic shift in kinase activity, and by a translocation of the protein itself (as revealed by immunoblotting; Fig. 2c) from the supernatant to the membrane. Similarly, geranylgeranylated rhodopsin kinase concentrates in the membrane fractions, whereas the non-isoprenylated form remains in the supernatant. But, neither the geranylgeranylated nor non-isoprenylated forms of the kinase exhibit any significant light-induced movement,

demonstrating a marked distinction between farnesylation and geranylgeranylation in protein function, namely the light-dependent regulation of rhodopsin kinase translocation (Fig. 2b, c).

The ability of the farnesylated kinase to translocate from supernatant to ROS membranes in a light-dependent manner represents a novel mechanism of enzyme compartmentalization. The light-independent association of geranylgeranylated rhodopsin kinase with ROS membranes suggests that the farnesylated kinase associates with activated rhodopsin, rather than solely through an isoprenoid lipid bilayer insertion. The farnesyl moiety seems to be critically involved in this interaction because non-isoprenylated rhodopsin kinase does not associate closely with dark-adapted or light-bleached ROS membranes (Fig. 2b). These data support the concept that the diminished activity of

Fig. 2 *a*, Relative activities of farnesylated (C15), geranylgeranylated (C20), and non-isoprenylated (C0) rhodopsin kinase as measured by phosphorylation of rhodopsin in ROS membranes. The bar graph shows the means $\pm$ s.e.m. of five separate experiments. Values have been normalized to the wild-type kinase activity (*) (23 pmol P_i incorporated per min per μ g kinase). The amount of phosphate incorporated over the time period examined corresponded to $<5\%$ of the rhodopsin concentration. *b*, Translocation-phosphorylation assay of C15, C20 and C0 rhodopsin kinase where S is the supernatant activity and M is the membrane activity as measured by ^{32}P incorporation into light-activated rhodopsin in ROS membranes⁵. The 'no kinase' lanes represents vector-transformed COS-7 supernatant activities. Data are representative of a typical experiment, replicated four times with comparable results. *c*, Immunoblot of translocation assay, as described in methods, except that instead of assaying for activity samples were electrophoresed on 7.5% SDS-polyacrylamide gels, transferred to nitrocellulose and analysed by standard western procedures using anti-rhodopsin kinase. METHODS. Samples of supernatants were immunoprecipitated with anti-rhodopsin kinase or anti- β ARK antibodies. Immunoprecipitated proteins were electrophoresed on 10% SDS-polyacrylamide gels and after autoradiography, corresponding bands were excised, rehydrated and digested at 37 °C for 48 h with Tris-buffered saline containing 100 μ g ml^{-1} proteinase K. Concentrations of the various kinases were determined relative to the wild-type kinase in each case by scintillation counting. Kinase assays were done using urea-treated ROS membranes as described^{1,5}. Assays were done in 75 mM Tris-HCl, pH 7.5, 10 mM Magnesium acetate 5 mM DTT, 20 μ M rhodopsin, and 100 μ M ATP at 25 °C for 10 min, quenched with 35 μ l SDS-PAGE loading buffer, and the phosphorylated rhodopsin separated from other proteins by electrophoresis on a 10% SDS-polyacrylamide gel. Radioactive protein bands corresponding to rhodopsin were excised from the dried gel and subjected to scintillation counting. Translocation assays were done in polycarbonate centrifuge tubes (Beckman 8 $\times$ 34 mm) each tube containing a total volume of 40 μ l translocation buffer (50 mM Na_2HPO_4 , pH 7.4, 100 mM NaCl, 1 mM $MgCl_2$, 5 mM DTT, 10 μ M protein kinase A inhibitor and protease inhibitors) and 20 μ M rhodopsin (urea-treated ROS membranes). Diluted kinase supernatants were added and one tube of each pair was kept in the dark and the other illuminated with fluorescent white light at 25 °C for 3 min.



Both tubes were then centrifuged at 350,000g for 5 min at 4 °C. Supernatants were quickly removed with 50 μ l Hamilton syringe under room light. Bleached rhodopsin was added to a final concentration of 20 μ M. Pellets were resuspended in translocation buffer plus one volume of ROS dilution buffer (5 mM Tris-HCl, pH 7.4, mM DTT, 2 mM $MgCl_2$, 65 mM NaCl) to equalize the sample volumes. Four samples (dark/supernatant, dark/membrane, light/supernatant, light/membrane) were assayed for kinase activity (as described previously) or immunoblotted.

the non-isoprenylated form is a function of an impaired farnesyl-dependent association of the kinase with the membrane-embedded receptor, and indicate that the receptor itself may act as a farnesyl docking site.

Light-activated rhodopsin serves as a substrate for the related kinase β ARK¹⁸. We thus examined whether the activation of β ARK by the geranylgeranylated $\beta\gamma$ dimer could be bypassed by directly isoprenylating β ARK with either the farnesyl or geranylgeranyl isoprenoid. The isoprenylated β ARKs exhibited up to a threefold enhancement in activity towards bleached rhodopsin in ROS membranes (Fig. 3a); an essentially identical enhancement in activity was observed for β_2 -adrenergic receptor phosphorylation (data not shown). We also found that isoprenylation of β ARK significantly increased ROS membrane association (Fig. 3b, c).

Both the activity and the membrane association of β ARK increase with integral isoprenylation of the enzyme, with non-isoprenylated β ARK being localized to the supernatant, farnesylated β ARK having an intermediate supernatant/membrane distribution and geranylgeranylated β ARK being fully membrane-associated (Fig. 3b, c). These results indicate that artificially targeting β ARK to the plasma membrane by direct isoprenylation mimics the enhancing effects of $\beta\gamma$ on enzyme activity. But no significant stimulus-dependent movement of any of the β ARK species was observed, showing that rhodopsin kinase may require sequence determinants in addition to isoprenylation to undergo light-dependent translocation.

In the case of β ARK, the stimulus dependence of translocation seems to be caused by the physical interaction of β ARK with $\beta\gamma$. Liberation of $\beta\gamma$ subunits from receptor-activated G proteins, which are presumably already membrane-associated through their own isoprenyl groups¹⁹, now provide membrane anchoring sites for β ARK. Thus, the unmasking of the $\beta\gamma$ anchor site for β ARK is totally dependent on receptor G-protein coupling. The fact that distinct molecular forms of $\beta\gamma$ subunits exist²⁰ suggests additional specificity of this process.

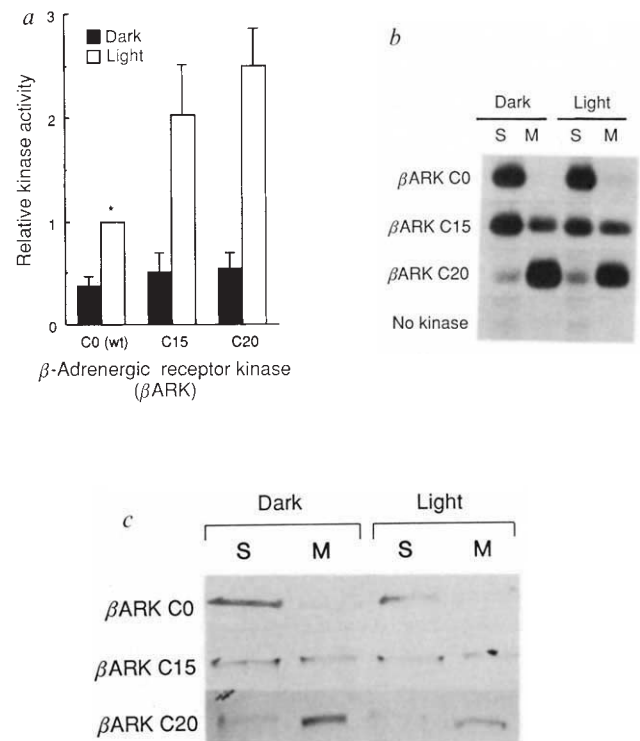


Fig. 3 *a*, Relative activities of C0, C15 and C20 β ARK as measured by phosphorylation of rhodopsin in ROS membranes. The bar graph shows the means $\pm$ s.e.m. of five separate experiments. Values have been normalized to the wild-type kinase activity (*) 1.8 pmol P_i incorporated per min per μ g kinase). *b*, Translocation-phosphorylation assay of C15, C20 and C0 β ARK with conditions as in Fig. 2. *c*, Immunoblot of β ARK species translocation as in Fig. 2c.

Isoprenylation, of G-protein-coupled receptor kinases, seems to provide a means for stimulus-orchestrated regulation of transmembrane signalling systems, by coupling cytosolic components with those of the plasma membrane. Effective regulation has been finely tuned by selection of the appropriate isoprenoid. For example, in the case of rhodopsin kinase the farnesyl group maintains the balance required to tip the kinase pool from cytoplasm to membrane at the command of a single photon. With β ARK, the regulation seems even more complex, presumably

in response to the need to coordinate signalling through numerous distinct receptor systems. This has been accomplished by the use of additional proteins, $\beta\gamma$, to carry the correct isoprenoid under direct control of the hormone/neurotransmitter-sensitive signalling system. Protein isoprenylation in translocation activation highlights yet another similarity between phototransduction and the hormone-stimulated signalling system. But precisely how each system uses this lipid modification to accomplish these similar functions is unique. □

Received 10 April; accepted 24 July 1992.

- Lorenz, W. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 8715–8719 (1991).
- Benovic, J. L., DeBlasi, A., Stone, W. C., Caron, M. G. & Lefkowitz, R. J. *Science* **246**, 235–240 (1989).
- Kuhn, H. *Biochemistry* **21**, 4389–4395 (1978).
- Strasser, R. H., Benovic, J. L., Caron, M. G. & Lefkowitz, R. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6362–6366 (1986).
- Inglese, J., Glickman, J. F., Lorenz, W., Caron, M. G. & Lefkowitz, R. J. *J. biol. Chem.* **267**, 1422–1425 (1992).
- Haga, K. & Haga, T. *J. biol. Chem.* **267**, 2222–2227 (1992).
- Mumby, S. M., Casey, P. J., Gilman, A. G., Gutowski, S. & Sternweis, P. C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5873–5877 (1990).
- Yamane, H. K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 5868–5872 (1990).
- Goldstein, J. L. & Brown, M. S. *Nature* **343**, 425–430 (1990).
- Glomset, J. A., Gelb, M. H. & Farnsworth, C. C. *Trends biochem. Sci.* **15**, 139–142 (1990).
- Hancock, J. F., Magee, A. I., Childs, J. E. & Marshall, C. J. *Cell* **57**, 1167–1177 (1989).
- Willumsen, B. M., Norris, K., Papageorge, A. G., Hubbert, N. L. & Lowy, D. R. *EMBO J.* **3**, 2581–2585 (1984).
- Jackson, J. H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 3042–3046 (1990).
- Casey, P. J., Solis, P. A., Der, D. C. & Buss, J. E. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8323–8327 (1989).

- Cox, A. D. & Der, C. J. *Crit. Rev. Oncogen.* (in the press).
- Pitcher, J. *et al.* *Science* (in the press).
- Mayor, F., Benovic, J. L., Caron, M. G. & Lefkowitz, R. J. *J. biol. Chem.* **262**, 6468–6471 (1987).
- Benovic, J. L., Mayor, F., Somers, R. L., Caron, M. G. & Lefkowitz, R. J. *Nature* **322**, 869–872 (1986).
- Simonds, W. F., Butrynski, J. E., Gautam, N., Unson, C. G. & Spiegel, A. M. *J. biol. Chem.* **266**, 5363–5366 (1991).
- Simon, M. I., Strathmann, M. P. & Gautam, N. *Science* **252**, 802–808 (1991).
- Casey, P. J., Thissen, J. A. & Moomaw, J. F. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8631–8635 (1991).
- Kinsella, B. T., Erdman, R. A. & Maltese, W. A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8934–8938 (1991).
- Hancock, J. F., Cadwallander, K., Paterson, H. & Marshall, C. J. *EMBO J.* **10**, 4033–4039 (1991).
- Andersson, S., Davis, D. L., Dahlback, H., Jorvall, H. & Russell, D. W. *J. biol. Chem.* **264**, 8222–8229 (1989).
- Farnsworth, C. C., Casey, P. J., Howald, W. N., Glomset, J. A. & Gelb, M. H. *Methods* **1**, 231–240 (1990).
- Arriza, J. L. *et al.* *J. Neurosci.* (in the press).

ACKNOWLEDGEMENTS. We thank J. F. Glickman, J. Pitcher, H. Attramadal and M. Hnatowich for assistance and discussions, J. L. Arriza for the β ARK antibodies and W. C. Stone and G. Irons for technical assistance.

Requirement of phosphatidylinositol 4,5-bisphosphate for α -actinin function

Kiyoko Fukami*, Kiyoshi Furuhashi†, Masaki Inagaki†, Takeshi Endo‡, Sadashi Hatano§ & Tadaomi Takenawa*||

* Department of Biosignal Research, Tokyo Metropolitan Institute of Gerontology, Sakae-cho, Itabashi-ku, Tokyo 173, Japan

† Laboratory of Experimental Radiology, Aichi Cancer Center Research Institute, Chikusa-ku, Nagoya, Aichi 464, Japan

‡ Department of Biology, Chiba University, Yayoi-cho, Chiba 260, Japan

§ Department of Molecular Biology, School of Science, Nagoya University, Chikusa-ku, Nagoya, Aichi 464, Japan

INOSITOL phospholipid turnover is enhanced during mitogenic stimulation of cells by growth factors¹ and the breakdown of phosphatidylinositol 4,5-bisphosphate (PtdInsP₂) may be important in triggering cell proliferation. PtdInsP₂ also binds actin-binding proteins to regulate their activity^{2–7}, but it is not yet understood how this control is achieved. The protein α -actinin from striated muscle contains large amounts of endogenous PtdInsP₂, whereas that from smooth muscle has only a little but will bind exogenously added PtdInsP₂. *In vitro* α -actinin binds to F-actin and will crosslink actin filaments, increasing the viscosity of F-actin solutions^{8,9}. We report here that α -actinin from striated muscle is an endogenous PtdInsP₂-bound protein and that the specific interaction between α -actinin and PtdInsP₂ regulates the F-actin-gelating activity of α -actinin. Although the F-actin-gelating activity of α -actinin from smooth muscle is much reduced compared with that from striated muscle, exogenous PtdInsP₂ can enhance the activity of smooth muscle α -actinin to the level seen in striated muscles. These results show that PtdInsP₂ is present in striated muscle α -actinin and that it is necessary for α -actinin to realize its maximum gelating activity.

As a first step towards elucidating the physiological significance of the interaction between PtdInsP₂ and cytoskeletal

proteins, we stained glycerol-treated chicken skeletal muscle with a monoclonal antibody that is specific for PtdInsP₂ (ref. 10). Although PtdInsP₂ is a component of the phospholipid membrane, it also stains densely in the Z-bands of the muscle sarcomere, where α -actinin is located¹¹ (Fig. 1). Antibody preabsorbed with PtdInsP₂ did not react with any structure in the myofibrils. These results indicate that endogenous PtdInsP₂ is probably binding to cytoskeletal proteins in the Z-bands of myofibrils.

To identify the PtdInsP₂-binding proteins in the Z-bands, we analysed myofibril proteins from striated muscle. Western blot analysis using anti-PtdInsP₂ antibody showed that the α -actinin component of striated muscle contained endogenous PtdInsP₂,

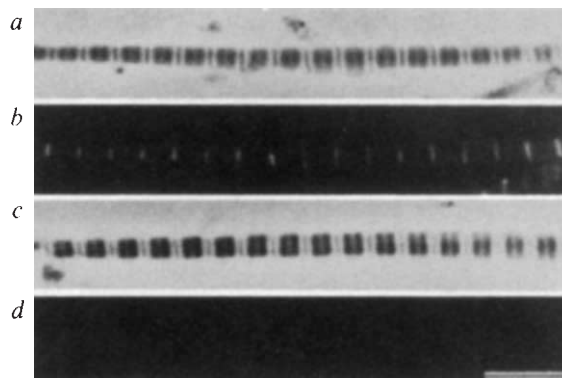


FIG. 1. Immunofluorescent localization of PtdInsP₂ in chicken skeletal muscle myofibrils. A myofibril was stained with the monoclonal antibody to PtdInsP₂. The phase-contrast and fluorescence micrographs of the myofibril are shown in a and b, respectively. A myofibril was treated with anti-PtdInsP₂ antibody preabsorbed with PtdInsP₂. The phase-contrast and fluorescence micrographs are shown in c and d. Scale bar, 10 μ m.

METHODS. Myofibrils were prepared from glycerol-extracted stretched pectoralis muscle of chicken as described¹², fixed on glass coverslips with 3.5% HCHO in PBS for 15 min and washed with PBS and 0.05% NaN₃. These myofibrils were then incubated with monoclonal antibody against PtdInsP₂ (ref. 2), and then fluorescein isothiocyanate-conjugated sheep anti-mouse IgG (affinity-purified, Cappel) (a and b). The myofibrils were also incubated with the antibody pretreated with PtdInsP₂, then with the fluorescein isothiocyanate-conjugated secondary antibody (c and d).

|| To whom correspondence should be addressed.

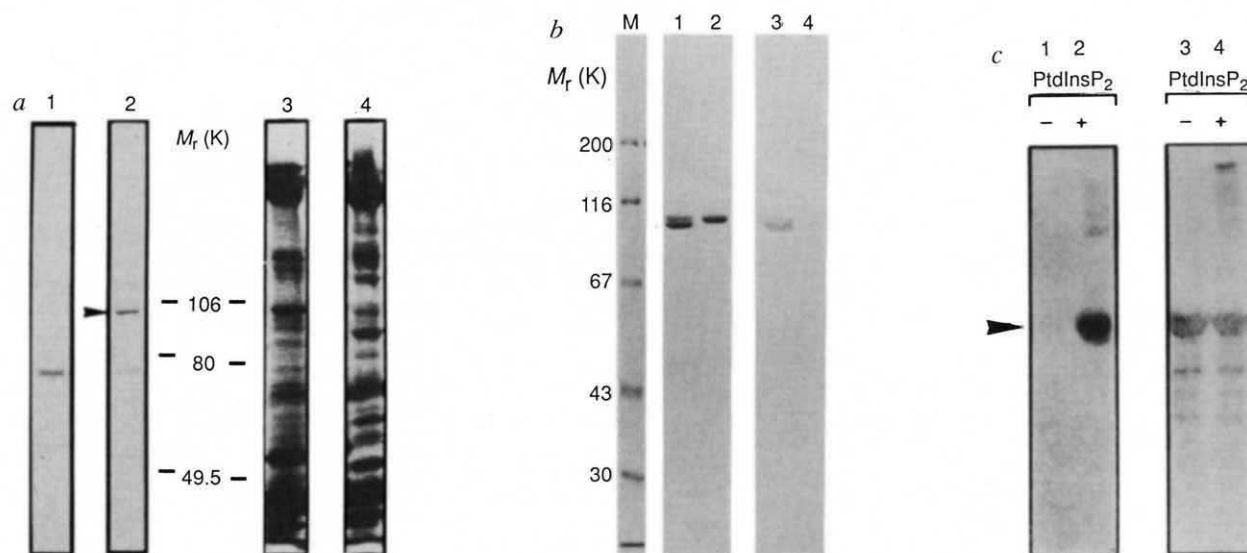


FIG. 2 Immunoblot analysis of α -actinin with anti-PtdInsP₂ monoclonal antibody. *a*, α -Actinin is a major PtdInsP₂-binding protein in striated muscle but not in smooth muscle. Total proteins of striated (lanes 1 and 2) and smooth (lanes 3 and 4) muscle were subjected to SDS-PAGE. Proteins were stained with Coomassie brilliant blue (lanes 3 and 4) or analysed by immunoblotting with anti-PtdInsP₂ antibody (lanes 1 and 2). *b*, Purified α -actinin from striated (lanes 1 and 3) and smooth muscle (lanes 2 and 4) was subjected to SDS-PAGE and the gel stained with Coomassie brilliant blue (lanes 1 and 2). Western blot with anti-PtdInsP₂ antibody was also done (lanes 3 and 4). Lane M, M_r standard marker proteins. Striated muscle α -actinin from chicken pectoralis appearing as a closely spaced doublet is attributed to different isoforms of striated muscle α -actinin¹⁹ (lane 1 and 3). *c*, Exogenously added PtdInsP₂ binds to smooth muscle α -actinin. Smooth muscle α -actinin incu-

bated with (lanes 2 and 4) or without (lanes 1 and 3) PtdInsP₂ was subjected to SDS-PAGE and western blot analysis with anti-PtdInsP₂ (lanes 1 and 2) or chicken gizzard α -actinin antibody²⁰ (lanes 3 and 4).

METHODS. *a*, Striated and smooth muscle were solubilized with SDS sample buffer and subjected to SDS-PAGE using a 7.5% separating gel according to Laemmli²¹. Immunostaining has been described²². *b*, Striated muscle and smooth muscle α -actinins were purified from chicken pectoralis muscle and gizzard as described²³. SDS-PAGE was done on a 10% polyacrylamide gel. *c*, PtdInsP₂ (620 μ M) and smooth muscle α -actinin (3.1 μ M) were mixed and incubated in phosphate-buffered saline at room temperature for 1 h and then SDS-PAGE on a 7.5% separating gel and immunoblotting were performed.

and that a 70K protein (M_r 70,000) was the principal PtdInsP₂-bound protein in smooth muscle (Fig. 2*a*); also, purified α -actinin from striated muscle was stained by anti-PtdInsP₂ antibody much more strongly than α -actinin from smooth muscle (Fig. 2*b*). Smooth muscle α -actinin bound strongly to exogenously added PtdInsP₂ (Fig. 2*c*). The binding of PtdInsP₂ to striated muscle α -actinin was so tight and stable that they remained bound even after electrophoresis on SDS-polyacrylamide gels.

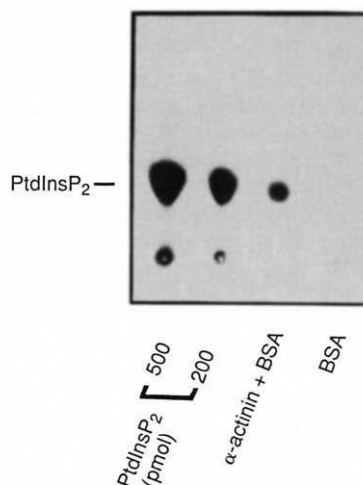
To confirm that the interaction between anti-PtdInsP₂ antibody and striated muscle α -actinin was due to endogenous PtdInsP₂ bound to α -actinin, we released PtdInsP₂ from α -actinin by extraction with chloroform and methanol (Fig. 3), although not all PtdInsP₂ could be removed by this method. By dot-blot analysis, the ratio of bound PtdInsP₂ was 20–30 mol to

1 mol α -actinin in striated muscle, but smooth muscle α -actinin contained less than 2 per cent of the PtdInsP₂ of striated muscle α -actinin.

To clarify the physiological relevance of the PtdInsP₂ association with striated α -actinin, the effects of inositol phospholipids on the gelating activity of smooth muscle α -actinin were measured by falling-ball viscometry, in parallel with striated muscles¹². α -Actinin purified from non-muscle cells and from smooth muscle was found to have gelating activity at low temperatures but not under physiological conditions^{13,14}. Striated muscle α -actinin formed a gel in the absence of inositol phospholipids, but smooth muscle α -actinin showed only weak gelating activity (Fig. 4*a*). Phosphatidylinositol has no effect on the gelating activity of α -actinin from either smooth muscle or striated muscle, and phosphatidylinositol 4-monophosphate has

FIG. 3 Identification of PtdInsP₂ extracted from purified striated α -actinin. Identification of lipids in purified α -actinin was carried out by thin-layer chromatography and immunostaining with anti-PtdInsP₂ antibody.

METHODS. Chicken striated muscle α -actinin (30 μ g) was treated with 1 ml 10% trichloroacetic acid after 100 μ g bovine serum albumin had been added as carrier protein, then kept at 4 °C for 1 h (α -actinin + BSA). Bovine serum albumin was similarly treated as a control (BSA). The mixture was centrifuged at 2,000*g* for 10 min. The precipitates were extracted with 2 ml chloroform/methanol (2/1, v/v) and 0.5 ml 1N HCl. The chloroform layers were evaporated under N₂ and the residue spotted onto a thin-layer chromatography plate with 500 and 200 pmol PtdInsP₂ as standards. The plate was developed in chloroform/methanol/NH₃/water (90/65/12/8, v/v) and immunostained with anti-PtdInsP₂ antibody¹⁰.



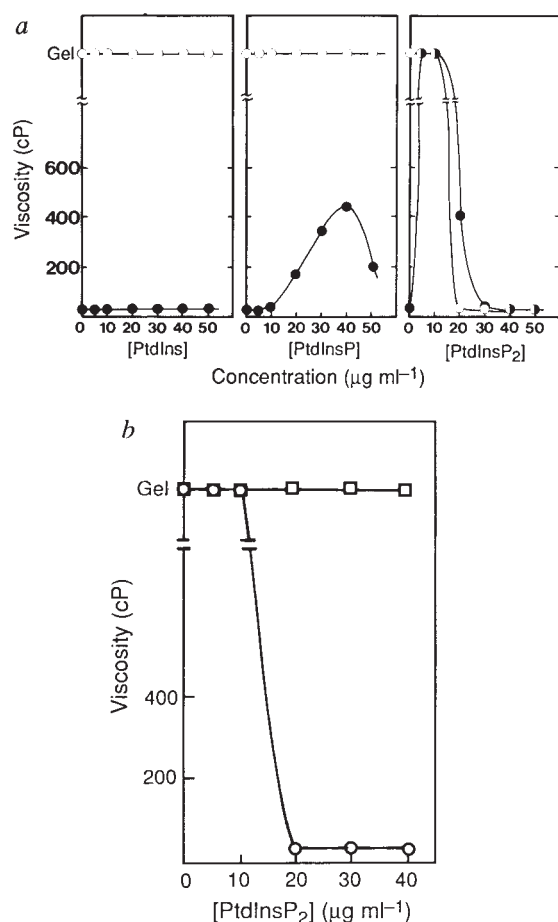


FIG. 4 Effects of exogenous inositol phospholipids on the gelating activity of α -actinin. *a*, the effects of PtdIns, PtdInsP and PtdInsP₂ on the interaction of striated muscle α -actinin (open circles) or smooth muscle α -actinin (filled circles) with F-actin were measured by falling-ball viscometry¹². *b*, The effect of PtdInsP₂ on gelating activity of striated muscle α -actinin was compared in the presence (squares) and absence (circles) of 0.5% Triton X-100. The gel represents the viscosity at which the ball does not fall. The concentrations shown are final concentrations in the reaction medium. The viscosity of actin alone and in the presence of smooth muscle α -actinin was 8 and 23 cP, respectively. Striated muscle α -actinin used in this experiment contained ~20–30 mol PtdInsP₂ in 1 mol α -actinin.

METHODS. Inositol phospholipids (30 μ l) in 50 mM Tris-HCl, pH 7.5, and α -actinin (60 μ l of 0.25 mg ml⁻¹) in buffer containing 10 mM Tris-HCl, pH 8.0, 0.1 mM EGTA, 10 mM β -mercaptoethanol and 0.1 mM PMSF were mixed and incubated at 25 °C for 1 h, after which 60 μ l 0.75 mg ml⁻¹ rabbit skeletal muscle G-actin in buffer containing 5 mM Tris-HCl, pH 7.5, 0.1 mM CaCl₂, 0.1 mM ATP and 0.2 mM DTT was added to the solution, followed by addition of KCl to 100 mM and MgCl₂ to 2 mM. Falling-ball viscometry was measured at 25 °C as described¹². All the manipulations in *b* were carried out in the presence of 0.5% Triton X-100.

only a slight activating effect. However, PtdInsP₂ has a marked effect on the gelating activity of smooth muscle α -actinin. This activity is increased to the level observed in striated muscle by 5–10 μ g ml⁻¹ of PtdInsP₂ (filled circles in Fig. 4*a*). A further increase in concentration of PtdInsP₂ gave a reduction in gelating activity of α -actinin to basal level. This inhibitory effect by a high concentration of PtdInsP₂ on α -actinin function was also seen in striated muscle. This phenomenon may be due to the formation of large PtdInsP₂ micelles that induce aggregation of the α -actinin-PtdInsP₂ complex and prevent crosslinking. This explanation is consistent with data indicating that gelating activity is retained maximally even in high concentrations of PtdInsP₂ if 0.5% Triton X-100 is present (Fig. 4*b*). None of the inositol phospholipids produced changes in the viscosity of F-actin. These results suggest that α -actinin, when bound to the optimum amount of PtdInsP₂, shows maximum F-actin-gelating activity and that this activity in striated α -actinin is higher than in smooth muscle α -actinin because of the endogenous binding of PtdInsP₂ to striated α -actinin. We have previously found that an F-actin-gelating protein from smooth muscle, filamin, is also regulated by PtdInsP₂ *in vitro*¹⁵. Phosphatidylinositol 4,5-bisphosphate inhibits the F-actin-gelating activity of filamin under conditions in which it increases crosslinking by α -actinin. As smooth muscle filamin does not bind endogenous PtdInsP₂ to any extent (Fig. 2*a*), the physiological significance of the regulation of filamin by PtdInsP₂ is not clear.

It has been proposed that α -actinin may anchor actin filaments to the plasma membrane, and may have the potential to bind lipids such as palmitic acid and diacylglycerol^{16,17}. We have shown that PtdInsP₂ is specifically localized in the Z-bands of striated muscle, that endogenous PtdInsP₂ binds to striated muscle α -actinin, and that the gelating activity of α -actinin is regulated by PtdInsP₂. These results indicate that α -actinin may

act as a multifunctional molecule through interaction with specific lipids. Whether the PtdInsP₂-induced modulation of α -actinin function is associated with the construction and maintenance of the network of the actin cytoskeleton, or whether the interaction between α -actinin and PtdInsP₂ is affected by extracellular stimuli such as epidermal and platelet-derived growth factors, which increase PtdInsP₂ turnover, are areas for further investigation. □

Received 22 May; accepted 15 July 1992.

1. Rozengurt, E. *Science* **234**, 161–166 (1986).
2. Lassing, I. & Lindberg, U. *Nature* **314**, 472–474 (1985).
3. Janmey, P. A. & Stossel, T. P. *Nature* **325**, 362–364 (1987).
4. Janmey, P. A., Iida, K., Yin, H. L. & Stossel, T. P. *J. Biol. Chem.* **262**, 12228–12236 (1987).
5. Yu, F., Johnston, P. A., Sudhof, T. C. & Yin, H. L. *Science* **250**, 1413–1415 (1990).
6. Heiss, S. H. & Cooper, J. A. *Biochemistry* **30**, 8753–8758 (1991).
7. Anderson, R. A. & Marchesi, V. T. *Nature* **318**, 295–298 (1985).
8. Ebashi, S., Ebashi, F. & Maruyama, K. *Nature* **203**, 645–646 (1964).
9. Maruyama, K. & Ebashi, S. *J. Biochem.* **58**, 7–12 (1965).
10. Matuoka, K., Fukami, K., Nakanishi, O., Kawai, S. & Takenawa, T. *Science* **239**, 640–643 (1988).
11. Masaki, T., Endo, M. & Ebashi, S. *J. Biochem.* **62**, 630–632 (1967).
12. Pollard, T. D. & Cooper, J. A. *Meth. Enzym.* **85**, 211–233 (1982).
13. Bennett, J. P., Zener, K. S. & Stossel, T. P. *Biochemistry* **23**, 5081–5086 (1984).
14. Landon, F., Gache, Y., Touthou, H. & Olomucki, A. *Eur. J. Biochem.* **153**, 231–237 (1985).
15. Furuhashi, K., Inagaki, M., Hatano, S., Fukami, K. & Takenawa, T. *Biochem. biophys. Res. Commun.* **184**, 121–126 (1992).
16. Meyer, R. K., Schindler, H. & Burger, M. M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4280–4284 (1982).
17. Burn, P., Rotman, A., Meyer, R. K. & Burger, M. M. *Nature* **314**, 469–472 (1985).
18. Endo, T. & Masaki, T. *J. Biochem.* **92**, 1457–1468 (1982).
19. Kobayashi, R., Itoh, H. & Tashima, Y. *Eur. J. Biochem.* **143**, 125–131 (1984).
20. Endo, T. & Masaki, T. *Biochem. biophys. Res. Commun.* **107**, 1467–1474 (1982).
21. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
22. Khyse-Anderson, J. *J. biochem. biophys. Meth.* **10**, 203–209 (1984).
23. Feramisco, J. R. & Burridge, K. *J. Biol. Chem.* **255**, 1194–1199 (1980).

ACKNOWLEDGEMENTS. We thank S. Tominaga and T. Takahashi for encouragement, M. Ohara for discussion, T. Obinata for microscope facilities, and S. Tokumasu for secretarial services. This research was supported in part by Grants-in-Aid for Scientific Research and Grants for Cancer Research from the Ministry of Education, Science and Culture of Japan and special coordination funds of the Science and Technology Agency of the Government of Japan.

Association between GTPase activators for Rho and Ras families

Jeffrey Settleman, Charles F. Albright,
Lauren C. Foster & Robert A. Weinberg

Whitehead Institute for Biomedical Research and Department of Biology,
Massachusetts Institute of Technology, Nine Cambridge Center,
Cambridge, Massachusetts 02142, USA

THE *ras*-related low-molecular-mass GTPases participate in signal transduction involving a variety of cellular functions, including cell-cycle progression, cellular differentiation, cytoskeletal organization, protein transport and secretion^{1,2}. The cycling of these proteins between GTP-bound and GDP-bound states is partially controlled by GTPase activating proteins (GAPs) which stimulate the intrinsic GTP-hydrolysing activity of specific GTPases¹⁻⁶. The *ras* GTPase-activating protein (Ras-GAP) forms a complex with a second protein, p190 (*M_r* 190,000), in growth-factor stimulated and tyrosine-kinase transformed cells^{7,8}. At its carboxy-terminal end, p190 contains a region that is conserved in the breakpoint cluster region, *n*-chimaerin, and Rho-GAP⁹. Each of these three proteins exhibits GAP activity for at least one member of the *rho* family of small GTPases¹⁰. We have tested recombinant p190 protein for GAP activity on GTPases of the *ras*, *rho* and *rab* families, and show here that p190 can function as a GAP specifically for members of the *rho* family. Consequently, the formation of a complex between Ras-GAP and p190 in growth-factor stimulated cells may allow the coupling of signalling pathways that involve *ras* and *rho* GTPases.

We have recently isolated complementary DNAs encoding a cellular protein, p190, that forms a stable complex with Ras-GAP⁹. As well as having sequence homology with GTPases and a reported transcriptional regulator, p190 shares sequences with several proteins that are believed to have GAP activities. To test for potential GAP activity of p190, we expressed the complete coding sequence of a rat p190 cDNA clone in insect cells using a recombinant baculovirus. After infection, large amounts of p190 were seen in total cellular protein lysates, compared with lysates from uninfected cells (Fig. 1a, lanes 1 and 2). The authenticity of this protein was confirmed by immunoprecipitation by anti-p190 antibodies from [³⁵S]methionine-labelled cells (Fig. 1b). A two-step purification strategy was used to purify

p190 to about 50% homogeneity for use in GAP assays (Fig. 1a, lane 3).

GAP assays were performed on 14 different purified small GTPases. The GTPases were bound to [³²P]GTP, incubated in the presence or absence of p190, and GTP hydrolysis was then measured in a filter-binding assay. Of the fourteen GTPases that were tested, five showed a significant increase in GTP hydrolysis in the presence of p190 (Fig. 2); RhoA, RhoB, Rac1, Rac2, and cdc42Hs retained 15%, 22%, 6%, 2% and 24%, respectively, of their initial radioactivity. All five of these are members of the *rho* family of small GTPases. None of the remaining nine GTPases, including seven members of the *ras* family (*H-ras*, *N-ras*, *K-ras4b*, *rap1a*, *rap2*, *ralA*, and *ralB*) and two members of the *rab* family (*rab3a* and *yp11*) were affected by p190. The apparent GAP activity of p190 exhibits time-dependence (Fig. 3), as well as temperature- and dose-dependence (data not shown), as would be expected for an enzymatic activity. When [³²P]GTP was bound to RhoB, p190 caused no loss of radioactivity in a filter-binding assay, and the GTP was converted to GDP as assayed by thin-layer chromatography (data not shown). Thus, p190 stimulates the intrinsic nucleotide-hydrolysing activity of the *rho* GTPases.

To test whether the GAP activity of p190 is analogous to that of Ras-GAP¹¹ and Rho-GAP¹⁰, we tested the ability of p190 to stimulate the GTPase of *rhoA* having a glycine to valine substitution at codon 14 (RhoA^{V14}). A comparable substitution in the *ras* GTPase severely reduces Ras-GAP-induced GTPase activity¹¹. Indeed, the GTPase of RhoA^{V14} was not stimulated by p190 (Fig. 2). As with the *ras* GTPase¹², we expect that RhoA^{V14} is still capable of binding to p190.

To demonstrate that the GAP activity of p190 is due to sequences common to the breakpoint cluster region, *n*-chimaerin, and Rho-GAP, a portion of the p190 cDNA encoding the carboxy-terminal (*M_r* 35K) was expressed in *Escherichia coli* as a fusion protein, and a bacterial lysate containing this protein was tested for GAP activity. This lysate exhibited significant GAP activity on RhoB (Fig. 3), while a lysate from *E. coli* lacking this protein showed no detectable GAP activity (data not shown). These results indicate that the conserved sequences enable these proteins to serve as GAPs for members of the *rho* GTPase family. Additionally, the p85 regulatory subunit of PI-3 kinase¹³ and a recently identified protein that binds to the SH3 domain of the *Abl* protein³⁰ share sequences with this same GAP region of p190, and would also be expected to exhibit GAP activity for the *rho* GTPases. We suggest that individual members of the *rho* GTPase family may be recognized by

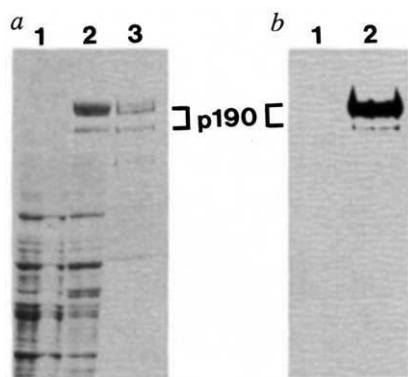


FIG. 1 Expression of p190 in insect cells. *a*, Coomassie stain of total cell lysate from uninfected (lane 1) and p190-baculovirus-infected (lane 2) insect cells, and (lane 3) partially purified p190 protein. *b*, Immunoprecipitation of p190 from uninfected (lane 1) and p190-baculovirus-infected (lane 2) insect cells metabolically labelled with [³⁵S]methionine.

METHODS. The complete p190 coding sequence was subcloned from rat cDNA 391 into the baculovirus expression vector pEV55. *Spodoptera frugiperda* sf9 insect cells were cotransfected with both the resulting plasmid and linearized wild-type baculovirus DNA (Invitrogen) and recombinant plaques were selected. The viral titer was amplified by propagation on *Trichoplusia ni* High Five cells (Invitrogen). Total cell lysate was prepared at 60 h post-infection by boiling the cell pellet in SDS-PAGE sample buffer, and proteins were resolved by 7.5% PAGE. For immunoprecipitation, p190-infected insect cells were metabolically labelled at 48 h post-infection for 5 h in methionine-free medium supplemented with 100 μ Ci ml⁻¹ [³⁵S]methionine (Translabel, ICN). Cells were lysed and immunoprecipitations with anti-p190 antibodies were performed⁹. Baculovirus-produced p190 protein was purified from infected High Five cells 60 h post-infection. Cells were lysed by sonication in 50 mM Tris, pH 7.5, 150 mM NaCl, 5 mM MgCl₂, 1 mM dithiothreitol (DTT), and 1 mM phenylmethylsulphonyl fluoride, the lysate was clarified by centrifugation at 60,000g, diluted in the same solution without NaCl to reduce the NaCl concentration to 50 mM, and proteins were resolved by MonoQ chromatography. Peak fractions containing p190 were precipitated with 40% ammonium sulphate and the pellet was redissolved in the original lysis buffer and desalted on a PD-10 column (Pharmacia).

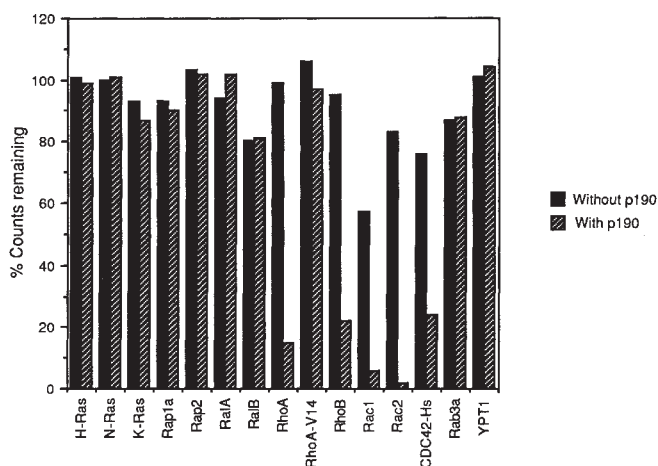


FIG. 2 GAP activity of p190 on various GTPases. Each GTPase was bound to [γ - 32 P]GTP, incubated with or without partially purified p190, and the radioactivity remaining on each GTPase after 10 min was determined in a filter-binding assay. The percentage of radioactivity remaining is relative to the initial radioactivity on each GTPase. Results for each GTPase are averages from at least two independent experiments.

METHODS. Low-molecular-mass GTPases were either obtained in purified form (rac1, rac2, rhoA, and cdc42Hs (ref. 17)) or purified from *E. coli* which express the GTPase. *E. coli* expression vectors were obtained for H-ras, N-ras, rap1a, rap2, ralA, rhoB, ypt1, and rab3a (refs 18–25). Expression vectors were constructed for K-ras4b by cloning of a polymerase chain reaction (PCR) fragment from pZnK (ref. 26) into pET-11a and cloning of a PCR fragment for ralB (ref. 27) into pET-5c. Most *E. coli*-produced proteins were purified by Q-sepharose chromatography and gel filtration on AcA54 columns (Pharmacia)²⁸. RhoA^{V14} was purified from *E. coli* as a glutathione S-transferase fusion protein using glutathione-sepharose (Sigma). The GTPase domain was released by cleavage with thrombin (Sigma) while the fusion protein was still bound to the beads²⁹. The purified GTPases (0.5–5 μ g of each) were bound to [γ - 32 P]GTP by incubation for 10 min at 25 °C in a solution containing 50 mM Tris, pH 7.5, 50 mM NaCl, 5 mM EDTA, 1 mg ml⁻¹ BSA, 1 mM DTT, 1 μ M [γ - 32 P]GTP, followed by the addition of MgCl₂ to a final concentration of 10 mM. GTPases were then incubated in 100 μ l of a solution containing 50 mM Tris, pH 7.5, 10 mM MgCl₂, 1 mg ml⁻¹ BSA, 1 mM DTT, with or without 100 ng p190 protein for 10 min at 25 °C. Samples were brought to 500 μ l with 50 mM Tris, pH 7.5, 10 mM MgCl₂, 1 mM DTT, and filtered through nitrocellulose (BA85, S&S). Filters were washed with 10 ml of the same solution and radioactivity remaining on the filters was determined.

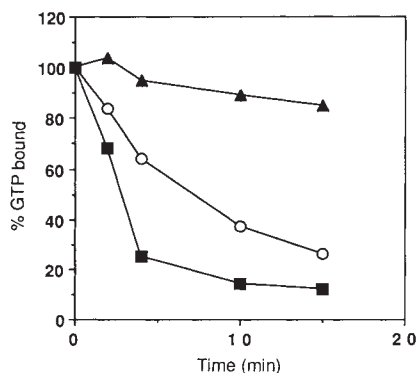


FIG. 3 GAP activity of p190 and C-terminal domain towards rhoB. The hydrolysis of GTP by RhoB was assayed in the absence of protein ($\blacktriangle$), or in the presence of p190 ($\blacksquare$) or the C-terminal fragment of p190 ($\circ$). **METHODS.** The C-terminal domain of p190 was cloned as a PCR fragment into the pET-5a expression vector (Novagen). Bacterial lysates from *E. coli* expressing the protein were prepared by sonication in 50 mM Tris, pH 7.5, 5 mM EDTA, and 3 mg ml⁻¹ lysozyme followed by centrifugation to remove insoluble material.

multiple GAP proteins, as is the case for the ras GTPases.

The formation of a complex between Ras-GAP and p190 following growth-factor stimulation may provide a mechanism to couple signals mediated by the ras and rho GTPases. Whereas the role of the ras proteins in transducing signals for mitogenesis and differentiation is well established^{1,2}, the role of the rho proteins is less clear, but it seems that these proteins regulate the actin cytoskeleton^{14,15} and cell polarity¹⁶. Thus, coupling of these pathways may coordinate progression through the cell cycle with reorganization of the cytoskeleton. $\square$

Received 15 June; accepted 27 July 1992.

- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125–132 (1990).
- Hall, A. *Science* **249**, 635–640 (1990).
- Evans, T., Hart, M. J. & Cerione, R. A. *Curr. Opin. Cell Biol.* **3**, 185–191 (1991).
- Lowy, D. R., Zhang, K., DeClue, J. E. & Willumsen, B. M. *Trends Genet.* **7**, 346–351 (1991).
- Downward, J. *Curr. Opin. Genet. Dev.* **2**, 13–18 (1992).
- Fry, M. J. *Curr. Biol.* **7**, 78–80 (1992).
- Ellis, C., Moran, M., McCormick, F. & Pawson, T. *Nature* **343**, 377–381 (1990).
- Moran, M., Polakis, P., McCormick, F., Pawson, T. & Ellis, C. *Molec. cell. Biol.* **11**, 1804–1812 (1991).
- Settleman, J., Narasimhan, V., Foster, L. C. & Weinberg, R. A. *Cell* **69**, 539–549 (1992).
- Diekmann, D. et al. *Nature* **351**, 400–402 (1991).
- Trahey, M. & McCormick, F. *Science* **242**, 1697–1700 (1987).
- Vogel, U. S. et al. *Nature* **335**, 90–93 (1988).
- Otsu, M. et al. *Cell* **65**, 91–104 (1991).
- Paterson, H. F. et al. *J. Cell Biol.* **111**, 1001–1007 (1990).
- Chardin, P. et al. *EMBO J.* **8**, 1087–1092 (1989).
- Adams, A. E. M., Johnson, D. J., Longnecker, R. M., Sloat, B. F. & Pringle, J. R. *J. Cell Biol.* **111**, 131–142 (1990).
- Hart, M. J., Polakis, P. G., Evans, T. & Cerione, R. A. *J. Biol. Chem.* **265**, 5990–5001 (1990).
- Gross, M. et al. *Molec. cell. Biol.* **5**, 1015–1024 (1985).
- Trahey, M. et al. *Molec. cell. Biol.* **7**, 541–544 (1987).
- Frech, M. et al. *Science* **249**, 169–171 (1990).
- Lerosey, I., Chardin, P., deGuzenberg, J. & Tavittian, A. *J. Biol. Chem.* **266**, 4315–4321 (1991).
- Frech, M., Schlichting, I., Wittinghofer, A. & Chardin, P. *J. Biol. Chem.* **265**, 6353–6359 (1990).
- Garrett, M. D., Self, A. J., van Oers, C. & Hall, A. *J. Biol. Chem.* **264**, 10–13 (1989).
- Baker, D., Wuestehube, I., Schekman, R., Botstein, D. & Segev, N. *Proc. natn. Acad. Sci. U.S.A.* **87**, 355–359 (1990).
- Zahraoui, A., Touchet, N., Chardin, P. & Tavittian, A. *J. Biol. Chem.* **264**, 12394–12401 (1989).
- McCoy, M. S., Bargman, C. I. & Weinberg, R. A. *Molec. cell. Biol.* **4**, 1577–1582 (1984).
- Chardin, P. & Tavittian, A. *Nucleic. Acids Res.* **17**, 4380 (1989).
- Tucker, J. et al. *EMBO J.* **5**, 1351–1358 (1986).
- Smith, D. B. & Johnson, K. S. *Gene* **67**, 31–40 (1988).
- Cicchetti, P., Mayer, B. J., Thiel, G. & Baltimore, D. *Science* **257**, 803–806 (1992).

ACKNOWLEDGEMENTS. We thank U. G. Knaus, X. M. Xu, G. Bokoch, A. Hall, R. Cerione, A. Wittinghofer, I. Lerosey, J. de Gunzburg, A. Tavittian, P. Chardin and N. Segev for providing cDNAs, purified proteins and plasmids; V. Narasimhan for help in construction of the baculovirus vector and in procuring GTPases and S. Egan, A. Sizeland, and D. Shaywitz for comments on the manuscript. J.S. and C.A. were supported by the American Cancer Society. This work was supported by a grant from the U.S. National Cancer Institute.

A meiotic gene conversion gradient opposite to the direction of transcription

Robert E. Malone, Steven Bullard, Steven Lundquist, Sangkyu Kim & Trudee Tarkowski

Department of Biology, University of Iowa, Iowa City, Iowa 52242, USA

GENETIC recombination involves classical crossing-over and gene conversion (aberrant segregation). In fungi that produce an ascus containing four spores, a gene conversion event is manifested as 3:1 or 1:3 (or more rarely 4:0 or 0:4) segregations, in contrast to the normal mendelian 2:2 segregation^{1,2}. Polarity is one of the properties of gene conversion; in almost all cases the frequency of conversion exhibits a gradient across the gene monitored¹. The frequency of conversion is usually independent of the specific allele used as a marker, but dependent on its location. An interpretation of conversion polarity is that it is caused by the existence of specific initiation sites for meiotic recombination, located at the high end of the polarity gradient. Here we show that the polarity gradient for the *HIS2* gene of *Saccharomyces cerevisiae* is high at the 3' end of the gene, implying that the promoter of *HIS2* is not the initiation site.

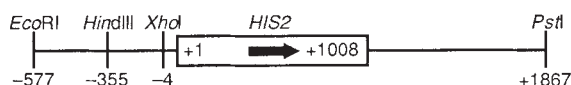


FIG. 1 Map of the *HIS2* gene. The coding region is indicated by the open box, the first base of which is designated as +1. The arrow indicates the direction of transcription. The TATA box is located at position -95, within the *HindIII*-*XhoI* fragment.

In *S. cerevisiae*, regions at the 5' end of genes are associated with a high frequency of conversion, hence polarity gradients are found in the same direction as transcription, in the *ARG4* gene^{1,3} (9.1% at the 5' end, 0.4% near the end of the coding region), in the *DED81* gene (15% at the 5' end, 4% at the 3' end)⁴ and in the *HIS4* gene⁵ (in a yeast strain that exhibits a high frequency of recombination). Furthermore, deletions in the promoter regions of *ARG4* and *HIS4* lead to reduced gene conversion^{5,6}, whereas a deletion of the TATA sequence of the *HIS4* gene that leaves the upstream activator sequences intact has no significant effect on gene conversion⁷. These studies indicate that meiotic recombination initiation sites may be related to promoter regions, although not necessarily to transcription *per se*^{7,8}.

The gene *HIS2* is reported to have a high frequency of conversion (14–18%)⁹. We cloned (Fig. 1) and sequenced (R.E.M., S.B., S.L., S.K. and T.T., manuscript in preparation) the gene and used the clone to make mutations *in vitro*. We placed the mutations in the genome by gene-replacement techniques^{10,11}. Each mutation was then crossed with a wild-type (*HIS2*) strain, allowed to go through sporulation and meiosis, and the resulting spores were genetically analysed. Eight different mutations have been analysed for conversion, and the results are shown in Fig. 2 and Table 1. Our results show that the polarity gradient for conversion at the *HIS2* gene is opposite to the direction of transcription. This indicates that the promoter region of the *HIS2* gene is unlikely to be associated with the initiation of recombination in this region.

We analysed recombination in the absence of the *HIS2* promoter region to provide support for this interpretation (Table 1). The diploid homozygous for this 351 base pair (bp) deletion, was then analysed for gene conversion. Rather than decreasing, the frequency actually increased from 11.7 to 20.1%. We conclude that the promoter region for the *HIS2* gene does not seem to be required for the initiation of meiotic recombination at this locus. There is a poly(dA . dT) sequence¹² in the *HIS2* promoter

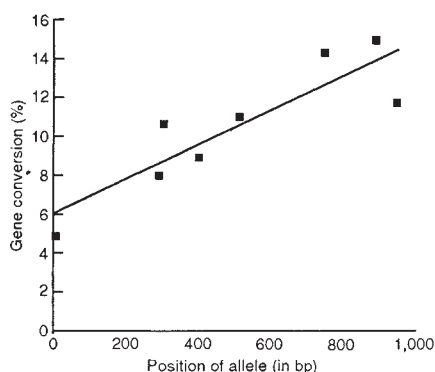


FIG. 2 Polarity of gene conversion at the *HIS2* locus. Each mutation analysed was created *in vitro* and then used to replace the wild-type gene by two-step gene replacement^{10,11} in strain RM96-15A. This strain was then crossed with RM182-55C, a *HIS2* strain of opposite mating type. The resulting heterozygous diploids were then sporulated and dissected, and the spores genetically tested as described⁹ for the *ARG4* gene. The least-squares line is shown, described by the equation $y = 0.0088x + 6.006$ ($r^2 = 0.776$).

TABLE 1 Gene conversion data for the *HIS2* gene

Allele	Location (bp)	Tetrads dissected	Conversion events				Gene conversion (%)
			3:1	1:3	4:0	0:4	
<i>his2-NSI</i>	+6	326	8	8	0	0	4.9
<i>his2-2</i>	+290	312	14	11	0	0	8.0
<i>his2-BGL</i>	+301	330	22	13	0	0	10.6
<i>his2-bam</i>	+402	336	15	15	0	0	8.9
<i>his2-bcl</i>	+509	303	12	19	2	0	10.9
<i>his2-xho</i>	+749	391	25	26	3	1	14.1
<i>his2-acc</i>	+885	308	23	22	0	1	14.9
<i>his2-hpa</i>	+947	377	22	20	1	1	11.7
<i>his2-hpa delHX</i>		214	18	25	0	0	20.1
<i>HIS2 delHX</i>							

Conversion data for all alleles analysed at the *HIS2* locus. The mutations in the *HIS2* gene were constructed as follows: *his2-NSI*, a single base insertion created by oligo-directed *in vitro* mutagenesis; *his2-2*, a single base substitution created by ultraviolet mutagenesis; *his2-BGL*, a single base substitution created by oligo-directed mutagenesis; *his2-bam*, a four-base insertion created by filling in a *Bam*HI site; *his2-bcl*, a four-base insertion created by filling in a *Bcl*I site; *his2-xho*, a four-base insertion created by filling in a *Xho*I site; *his2-acc*, a two-base insertion created by filling in an *Acc*I site; and *his2-hpa* created by inserting an 8 bp *Cla*I linker in a *Hpa*I site. The details of the construction of each mutation will be described (R.E.M., S.B., S.L., S.K. and T.T., manuscript in preparation). No significant deviations from parity (equal proportions of the 3:1 and 1:3 tetrads²) were observed in any of the crosses. The *HIS2* promoter is located between the *Hind*III (at -355) and the *Xho*I site (at -4) because a clone of the *HIS2* region, extending from the upstream *Hind*III site to the *Pst*I site located 860 bp downstream from the end of the coding region (see Fig. 1), expresses the *HIS2* gene and confers a His⁺ phenotype on a strain containing a complete deletion of the chromosomal *HIS2* copy. The 351-bp *Hind*III-*Xho*I fragment was deleted *in vitro* and the deletion placed in the chromosome^{10,11} of the *HIS2* strain and the *his2-hpa* strain; the resulting diploid was then analysed for conversion.

at positions -151 to -172. As deletion of the promoter did not reduce conversion at *HIS2*, this poly(dA . dT) sequence, unlike the one in *ARG4*, is apparently not required for the initiation of meiotic recombination. One possibility that would explain the observed polarity gradient and maintain a relationship to transcription is a promoter, located on the 3' side of the *HIS2* gene, that causes transcription to occur in a 3' to 5' direction relative to the coding region, though this is unlikely. Mitotic and meiotic northern blot analysis with strand-specific probes have revealed no trace of messenger RNA being made from the wrong strand of the *HIS2* gene (data not shown).

Most genetic observations indicate that meiotic recombination is initiated at specific sites. Although the initiation sites for three genes in yeast seem to be the 5' end of the genes, and the polarity gradients in those genes therefore have the same direction as transcription, data from the *HIS2* locus suggest that this is not an absolute requirement for meiotic recombination. It is possible that the *HIS2* initiation site is associated with the promoter of a neighbouring gene, although we have not found any putative promoter sequence in the 860 bp downstream of the coding region (to the *Pst*I site). Even if the initiation site is found to be associated with the promoter region of an unknown downstream gene, our analysis of *HIS2* indicates that polarity gradients in *S. cerevisiae* can exist in either orientation relative to the direction of transcription, and, for this gene, are not dependent upon the promoter. □

Received 7 April; accepted 31 July 1992.

1. Fogel, S., Mortimer, R., Lusnak, K. & Tavares, F. *Cold Spring Harb. Symp. quant. Biol.* **43**, 1325–1341 (1979).
2. Petes, T., Malone, R. E. & Symington, L. in *The Molecular Biology of the Yeast Saccharomyces* (eds Broach, J., Pringle, J. & Jones, E.) 407–522 (Cold Spring Harbor Laboratory, New York, 1991).
3. Nicolas, A., Treco, D., Schultes, N. P. & Szostak, J. W. *Nature* **338**, 35–39 (1989).
4. Schultes, N. P. & Szostak, J. W. *Genetics* **126**, 813–822 (1990).
5. White, M. A., Wierdl, M., Detloff, P. & Petes, T. D. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9755–9759 (1991).
6. Schultes, N. P. & Szostak, J. W. *Molec. cell. Biol.* **11**, 322–328 (1991).
7. White, M., Detloff, P., Strand, M. & Petes, T. D. *Curr. Genet.* **21**, 109–116 (1992).
8. Schultes, N. P. & Szostak, J. W. *Genetics* **126**, 813–822 (1990).
9. DiCaprio, L. & Hastings, P. J. *Genetics* **84**, 697–721 (1976).
10. Rothstein, R. *Meth. Enzym.* **194**, 281–301 (1991).
11. Scherer, S. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4951–4955 (1979).
12. Struhl, D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8419–8423 (1985).

Loss of gene function through rapid mitotic cycles in the *Drosophila* embryo

Mike Rothe*, Michael Pehl, Heike Taubert & Herbert Jäckle

Max Planck Institut für biophysikalische Chemie, Abteilung Molekulare Entwicklungsbiologie, Am Faßberg, Postfach 2841, W-3400 Göttingen, Germany

* Present address: Molecular Biology Department, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, USA

THE early developmental period in *Drosophila* is characterized by rapid mitotic divisions, when the body pattern becomes organized by a cascade of segmentation gene activity^{1,2}. During this process localized expression of the gap gene *knirps* (*kni*) is required to establish abdomen segmentation^{3,4}. The *knirps*-related gene (*knrl*) encodes a *kni*-homologous nuclear hormone receptor-like protein^{5,6} and shares the spatial patterns of *kni* expression. The two genes differ with respect to the size of their transcription units; *kni* contains 1 kilobase and *knrl* 19 kilobases of intron sequences. The consequence of this difference in intron size is that *knrl* cannot substitute for *kni* segmentation function, although it gains this ability when expressed from an intronless transgene. Here we show that the length of mitotic cycles provides a physiological barrier to transcript size, and is therefore a significant factor in controlling developmental gene activity during short 'phenocritical' periods. The required coordination of cycle length and gene size provides severe constraints towards the evolution of rapid development.

The genes *kni* and *knrl* encode proteins that contain an almost identical 93-amino-acid amino-terminal domain⁵ characteristic of the DNA-binding motif of ligand-dependent vertebrate nuclear hormone receptors^{7,8}. The two genes map to the same location in the genome^{5,6} and their spatial expression patterns coincide precisely throughout embryogenesis⁶ (Fig. 1). These common features indicate that *knrl* carries a *kni*-like function⁶. In the absence of *kni* activity, embryos fail to develop most of the abdominal segments (Fig. 2a). This defective phenotype was also observed in embryos that lack both *kni* and *knrl* gene

activities^{6,9} (Fig. 2b). If such embryos contain a *kni* DNA transgene construct, the abdominal defect is rescued into a wild-type pattern (Fig. 2c, d). Thus, *knrl* activity is not required to establish abdominal segments in the wild-type embryo.

But the *knrl* protein (Knrl) has the potential to function like the *kni* protein (Kni) both *in vitro* and *in vivo*. Knrl can bind *in vitro* to Kni target sequences and can act as a transcriptional repressor that antagonizes transcriptional activation by the fly morphogen *bicoid* (*bcd*) (ref. 10) in a manner indistinguishable from Kni (Fig. 3). In the embryo, ectopic expression of Knrl from a complementary DNA-derived *knrl* minigene leads to the same 'anti-*kni*' phenotype observed after ectopic expression of Kni (Fig. 2e and f). Thus, Knrl is also likely to interact *in vivo* with the same DNA target as Kni. To test this proposal, we examined the *kni* rescuing activity derived from a modified *kni* rescue transgene in which the sequences of the DNA-binding domain of Kni are replaced by the corresponding Knrl sequence (Fig. 2k). Activity from this *kni/knrl* hybrid gene restores the abdominal segment pattern in *kni*-deficient embryos. Furthermore, the replacement of the complete Kni coding sequence by a cDNA-derived Knrl minigene results in a dosage-dependent development of abdominal segments (Fig. 2g-i). Thus, Knrl also has a Kni-like biological function corresponding to the activity observed in some weak *kni* mutant alleles³ (Fig. 2j), but *knrl* still fails to provide segmentation function in wild-type embryos.

It is the size of the *knrl* transcription unit that prevents the *knrl* gene from participating in the process of abdominal segmentation. The *knrl* primary transcript has about 23 kilobases (kb), whereas *kni* is only 3 kb in size⁴ yielding different intron sizes (Fig. 3a). Expression of both *kni* and *knrl* is observed during mitotic cycle 13 when the preblastoderm embryo consists of a peripheral monolayer of nuclei derived from 12 rapid and synchronous mitotic cycles in the absence of cell divisions¹¹. By means of *in situ* hybridization to whole-mount embryos, *knrl* transcripts were found to accumulate in two distinct spots of the nuclei. This subnuclear pattern was observed with a probe directed against the 5' end of the transcript (Fig. 4a), but not with a 3'-probe (Fig. 4c). In contrast, *kni* as well as *knrl* transcripts derived from the intronless *knrl* transgene were detected in the cytoplasm surrounding the syncytial nuclei with both 5'- and 3'-probes (Fig. 4b, d). During the expanded mitotic cycle 14 (refs 12, 13), in which nuclei become separated by ingrowing

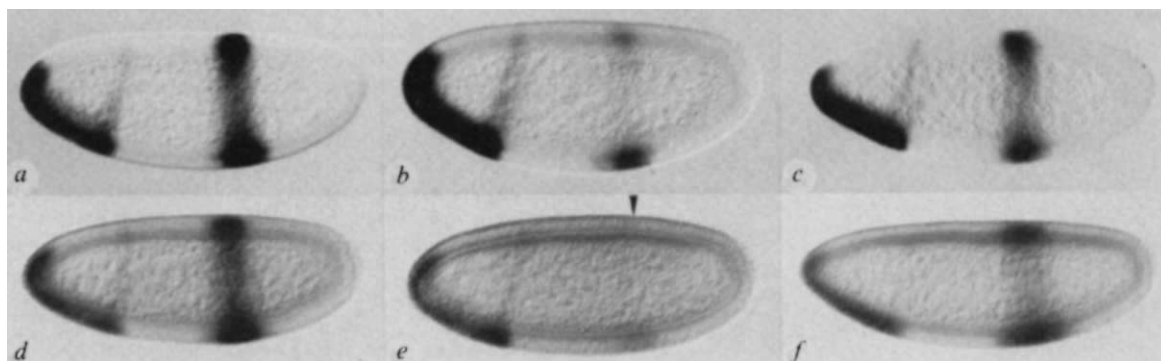
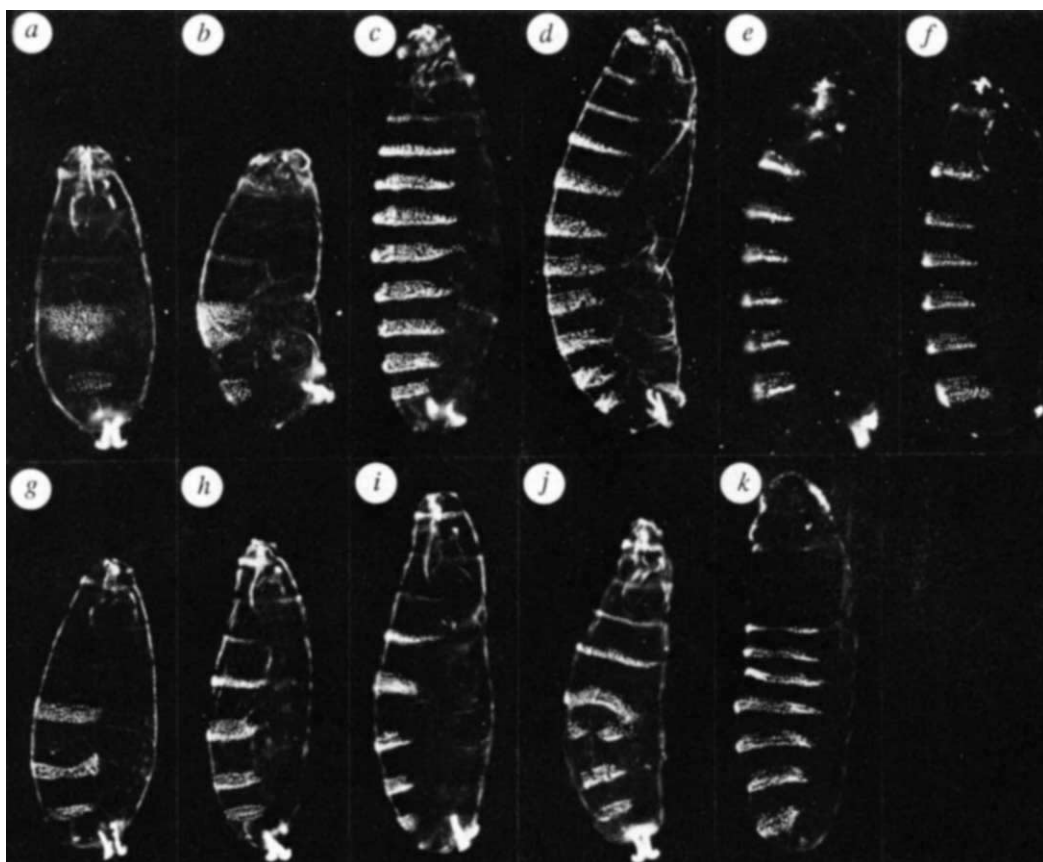


FIG. 1 Blastoderm expression of *kni* and *knrl*. Whole-mount *in situ* hybridizations (a-c) and antibody stainings (d-f) of embryos during nuclear cycle 14 are shown. a, b Wild-type embryos hybridized with *kni* cDNA⁴ (a) and *knrl* cDNA⁵ (b). c, Homozygous P[w⁺; *kni*5'*knrl*cDNA] embryo hybridized with *knrl* cDNA. Note the identical spatial expression patterns (but different intensities) in the prospective abdominal and head regions^{5,6}. d-f, Expression of *kni* (d)¹⁵ and *knrl* protein (e, f) in wild-type (d, e) and homozygous P[w⁺; *kni*5'*knrl* cDNA] embryos (f). The protein patterns in d and f are indistinguishable, and both proteins can already be observed during nuclear cycle

13 (data not shown). In contrast, *knrl* protein derived from the endogenous wild-type gene (e) can be detected no earlier than advanced nuclear cycle 14. Arrowhead in e marks the posterior *knrl* protein domain. Anterior pole of embryos is left, dorsal side up.

METHODS. *In situ* hybridization was done according to a modified standard protocol^{16,17}. Monospecific rabbit anti-*knrl* antibodies were generated against the bacterially expressed *knrl* carboxy-terminal domain (nucleotides 971-1,358) as described in ref. 15, and antibody stainings were done as described in ref. 18.

FIG. 2 Rescue of mutant *kni* segmentation phenotype by *kni* and *knrl* transgene activity. *a*, Larval cuticle preparations of *kni*^{FC13} (ref. 3) and *b*, of *Df(3L)ri*^{XT1} homozygous embryo¹⁹. Note identical abdominal segmentation phenotypes although the latter embryos are *kni/knrl* double mutants. *c*, Homozygous *kni*^{FC13} and *d*, homozygous *Df(3L)ri*^{XT1} embryos each of which carries two copies of the *kni* rescue construct *P[ry⁺;kni]*. In both embryos, the mutant abdominal segment pattern is rescued into a wild-type pattern, but only a weak rescue of head defects is seen in *kni/knrl* double mutant embryos (compare *b* and *d*). *e*, *f*, Wild-type embryos in which heat-shock-induced *hsp70-kni* (*e*) or *hsp70-knrl* (*f*) occurred, develop same 'anti-*kni*' phenotype (abdominal segments 7 and 8 are fused, and the thoracic segments, abdominal segment 1 and posterior terminal structures are either missing or at least partially deleted). *g-i*, Homozygous *kni*³⁰¹ embryos¹⁹ containing one (*g*), two (*h*) or three copies (*i*) of the *knrl* rescue construct *P[w⁺;kni5'/knrl cDNA]*. Expression of the *knrl* cDNA under the control of the *kni* promoter rescues the mutant phenotype into the development of three (*g*), four (*h*) and five (*i*) abdominal segments; embryos shown in *h* and *i* resemble *j*, the weak *kni* phenotype of the mutant *kni14F* embryo³. *k*, Homozygous *kni*³⁰¹ embryo harbouring one copy of the chimaeric rescue construct *P[w⁺;kniknrlfing]* expressing a *kni* protein that contains the *knrl* finger domain. To identify the *kni/knrl* mutant embryos among the progenies, we marked them by double mutant combinations with *tailless* (*tl*⁶) (ref. 20). The abdominal *kni* segmentation phenotype is rescued to leave the *tl* phenotype. This control was also used for scoring the rescue of embryos in *c-e*, *g-i* (data not shown). **METHODS.** The *P[ry⁺;kni]* construct contains a 9.5-kb genomic *Sall*-*HindIII* fragment which comprises the *kni* transcription unit and ~6.5 kb of flanking sequences²¹ cloned into the P-element vector Carnegie 20 (ref. 22). For *P[w⁺;kniknrlfing]*, the finger region of the *kni* gene was replaced by the corresponding region of the *knrl* gene using polymerase chain reaction



directed mutagenesis and inserted into the CaSpeR P-element vector²³. The *P[w⁺;kni5'/knrl cDNA]* construct contains the coding sequence of the *knrl* cDNA⁵ as a 2.1 kb *BanII*-*SphI* fragment which was inserted between the *Nrul* site of the untranslated leader of the *kni* gene^{4,21} and the *Drosophila* $\alpha 1$ -tubulin polyadenylation signal²⁴ in CaSpeR. For heat-shock-induced ectopic expression of *kni* and *knrl*, the *Drosophila* *hsp70* promoter²⁵ and the $\alpha 1$ -tubulin polyadenylation signal were cloned into CaSpeR (*pCaSpeRhspNot*). The 1.4 kb *Nrul*-*MaeIII* fragment of the *kni* cDNA and the 2.1 kb *BanII*-*SphI* fragment of the *knrl* cDNA⁵ were inserted into the single *NotI* cloning site of *pCaSpeRhspNot* after subcloning into Bluescript (Stratagene). P-element-mediated transformation of flies was done as described²⁶, and transgenes were crossed into the appropriate mutant backgrounds. Rescue effects were assayed in cuticle preparations²⁷. Heat-shock-treatment of embryos (37 °C) occurred at 120-150 min after egg deposition. For each experiment at least two independent transformant lines were tested.

membranes forming the cellular blastoderm, *kni* transcripts continue to accumulate in the cytoplasm. The *knrl* transcripts are restricted to the nuclei of embryos, and only appear in the cytoplasm after about 20 min (Fig. 4*e-h*).

The lag-time between nuclear and cytoplasmic localization of *knrl* transcripts reflects the time required to synthesize a 23 kb primary transcript. According to the previously determined transcription rate of 1.4 kb min⁻¹ (ref. 14), *knrl* and *kni* full-size transcription would take ~16 and <3 min, respectively. These calculations agree with the observation that *kni* primary transcripts are completed during the 10-11 min interphase of cycle 13 (refs 12, 13), whereas the synthesis of full-size *knrl* primary transcripts is interrupted by the short mitotic cycle, and nascent transcripts become aborted through progression of the cell cycle¹⁴. The relatively late initial appearance of *knrl* transcripts in the cytoplasm of cycle 14 and the time when the corresponding proteins can be detected by antibody staining in wild-type embryos (cycle 14; see Fig. 1) are consistent with this proposal. But when functional Knrl is expressed from the *knrl* cDNA

minigene, it can be detected as early as Knl (Fig. 4). Thus, the lag-time between functional cDNA-derived Knrl expression and non-functional wild-type Knl expression delimits a transient 'phenocritical' period for abdominal gap gene activities between cycle 13 and midstage cycle 14.

As nascent transcripts are aborted during mitosis, the time of the mitotic cycle provides an efficient physiological barrier that determines the maximum size of functional transcription units at a given mitotic interphase. This implies that regulatory genes acting in rapidly proliferating cells need to be compact in size, and that evolutionary events towards a rapid development of organisms require coordination between the length of cell cycles and the size of regulatory genes¹⁴. It is possible that the short *kni* gene originates from *knrl* or a *knrl*-like ancestral gene, whose function in evolutionary older insects (which develop through longer embryonic cell cycles) was adapted after gene duplication to the extremely rapid development and the short phenocritical period for segmentation in *Drosophila*. □

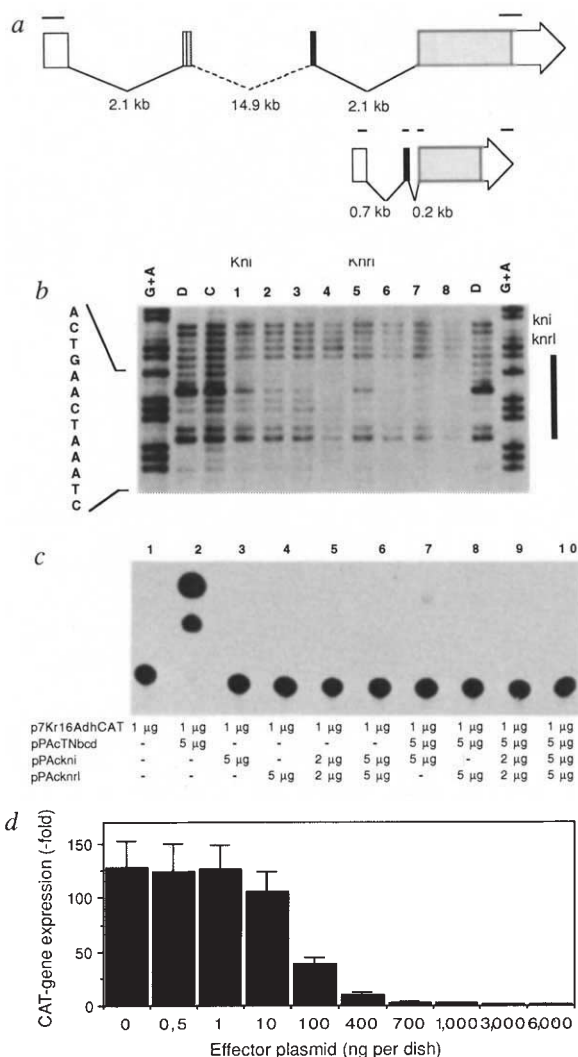
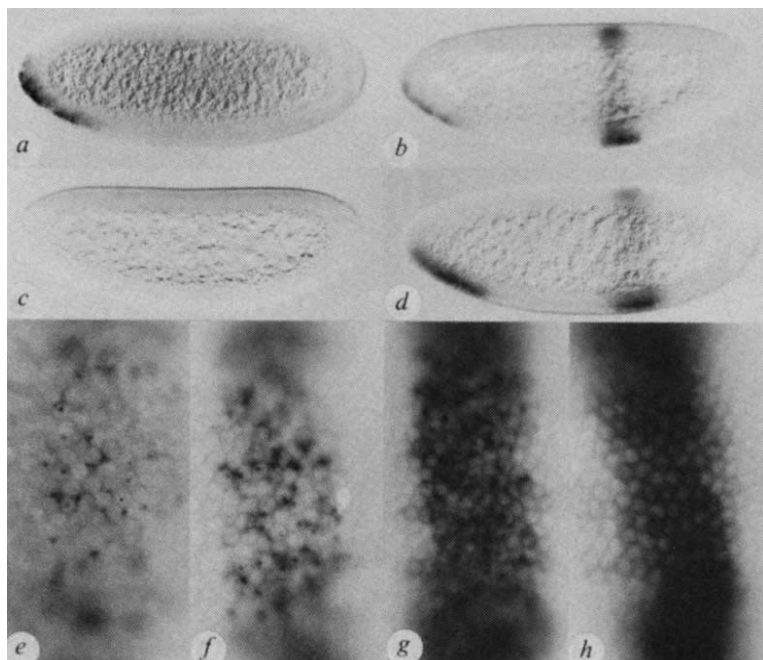


FIG. 3 The DNA-binding protein, Knrl, interacts through *kni* target sequences. **a**, Organization of the *knrl* (upper panel) and *kni* transcription unit (lower panel; ref. 4) as derived from the comparison of cDNA and genomic sequences. Because of the different numbers and sizes of introns, the *knrl* transcription unit comprises 23 kb of genomic DNA whereas the *kni* transcription unit comprises only 3 kb. Arrows indicate the direction of transcription; open bars the untranslated sequences; stippled bars are *kni*⁴ or *knrl*⁵ coding regions; and the dashed line indicates the 14.9 kb *knrl* intron. The exon encoding the identical first finger motif of Kni and Knrl^{5,6} is indicated as a filled bar. The positions of the *knrl* and *kni* 5' and 3' probes used for *in situ* hybridization are marked by a line above. **b**, DNA binding of Knrl and Kni to a 16-bp *cis*-acting sequence from the *Krüppel* (*Kr*) gene¹⁰ shown by *in vitro* footprinting bacterial extracts (2 µg, lanes 1 and 5; 4 µg, lanes 2 and 6; 6 µg, lanes 3 and 7; and 8 µg, lanes 4 and 8) containing Knrl (lanes 1–4) or Kni (lanes 5–8) were used¹⁰. Control lane C: footprinting reaction with 10-µg extracts from bacteria containing the T7 expression vector without *knrl* or *kni* coding sequences. Control lane D: footprinting reaction in the absence of bacterial extracts. G + A: Maxam–Gilbert sequencing reactions. Note that Knrl and Kni protect the same area (solid bar). **c**, Knrl and Kni act as transcriptional repressors antagonizing *bicoid* protein (BCD)-dependent gene activation in *Drosophila* Schneider tissue culture cells cotransfected with the indicated amounts of *bcd* (pPacTN-bcd) (ref. 28), *kni* (pPac-kni) (ref. 10) and *knrl* (pPac-knrl) effector plasmid DNA and the chloramphenicol acetyl transferase (CAT) reporter gene construct p7Kr16AdhCAT (ref. 10). Both Kni and Knrl repress *bcd*-dependent transcriptional activation mediated by the *Kr* 16-bp element. **d**, Cotransfection experiments with constant amounts of p7Kr16AdhCAT DNA (1 µg), pPacTN bcd effector DNA (3 µg), and increasing amounts of pPac-knrl effector plasmid. Bars represent the mean values from three independent experiments. Note that Bcd-dependent CAT gene activation decreases with the amount of *knrl* effector plasmid added in a fashion analogous to *kni*¹⁰. **METHODS.** Cloning of DNA and handling of bacteria were done by standard methods²⁹. For DNase I footprinting analysis, the DNA-binding region of Knrl (nucleotides 538–970) (ref. 5) was expressed by use of the bacterial expression vector pET3c system³⁰. Full-length Kni¹⁵ and Bcd³¹ were produced as described. Bacterial extracts, footprinting and Maxam–Gilbert reactions were done as described previously³¹. For expression of full-length Knrl in tissue culture cells the 2.1 kb *Bam*II–*Sph*I fragment of the *knrl* cDNA *imd2* (ref. 5) was inserted into the *Bam*HI site of the pPac vector³² by blunt-end ligation. Transfection of Schneider S2/M3 cells and CAT assays³³ were done as described¹⁰.

FIG. 4 Progression of *knrl* transcription during nuclear cycles 13 and 14. **a, c**, Wild-type embryos at nuclear cycle 13 hybridized with the *knrl* 5' (**a**) and 3' (**c**) probe. With the 5' probe a punctuated pattern of nascent nuclear *knrl* transcripts can be observed in the anterior domain (see also **f**). No transcripts can be detected with the 3' probe at this stage. **b, d**, homozygous P[*w*⁺; *kni*5'*knrl* cDNA] embryos at nuclear cycle 13 hybridized with the same *knrl* 5' (**b**) and 3' (**d**) probes. Cytoplasmic localization of transcripts derived from the *knrl* cDNA minigene can be observed with both probes. This pattern is identical to the *kni* expression pattern observed with 5'- and 3'-specific *kni* probes in wild-type embryos (data not shown). **e–h**, Posterior *knrl* expression domain in *knrl*^{FC13} embryos hybridized with the *knrl* 5' probe. Embryos are 5–10 min (**e**), 15–20 min (**f**), 25–30 min (**g**) and 45 min (**h**) into nuclear cycle 14. *knrl* transcripts initially appear as nuclear dots of nascent transcripts. After ~20 min, transcripts can be detected in the cytoplasm where they rapidly accumulate. Transcription of *knrl* ceases after ~30 min when nuclear transcripts can no longer be observed and all transcripts become localized in the cytoplasm. With the onset of gastrulation, *knrl* transcripts are degraded⁶. Essentially the same pattern is observed in wild-type embryos (data not shown). Anterior pole of embryos is left, dorsal up.

METHODS. The 5'- and 3'-specific *kni* and *knrl* probes comprise nucleotides 1–393 and 2,166–2,611 of the *knrl* cDNA⁵ and 202–491 and 1,799–2,046 of the *kni* cDNA⁴, respectively. For standardized *in situ* hybridizations, equal amounts of 5' and 3' probe of the same length were labelled in parallel reactions to be used on batches of pretreated embryos. Staining reactions were done in parallel (for details see Fig. 1 legend). Age determination of embryos was



based on the depth of membrane invagination³⁴ and morphological features^{11,12}.

Received 15 June; accepted 24 July 1992.

1. Akam, M. *Development* **101**, 1–22 (1987).
2. Ingham, P. *Nature* **335**, 25–34 (1988).
3. Lehmann, R. *Development* **104** (suppl.), 17–27 (1988).
4. Nauber, U. *et al. Nature* **336**, 489–492 (1988).
5. Oro, A. E. *et al. Nature* **336**, 493–496 (1988).
6. Rothe, M., Nauber, U. & Jäckle, H. *EMBO J.* **8**, 3087–3094 (1989).
7. Evans, R. M. *Science* **240**, 889–895 (1988).
8. Green, S. & Chambon, P. *Trends Genet.* **4**, 309–314 (1989).
9. Oro, A. E., Urnesono, K. & Evans, R. M. *Development* **107** (suppl.), 133–140 (1989).
10. Hoch, M., Gerwin, N., Taubert, H. & Jäckle, H. *Science* **256**, 94–97 (1992).
11. Campos-Ortega, J. A. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster* (Springer, Berlin, 1985).
12. Foe, V. E. & Alberts, B. J. *Cell Sci.* **61**, 31–70 (1983).
13. Foe, V. E. *Development* **107**, 1–22 (1989).
14. Shermoen, A. W. & O'Farrell, P. H. *Cell* **67**, 303–310 (1991).
15. Pankratz, M. J. *et al. Cell* **61**, 309–317 (1990).
16. Tautz, D. & Pfeifle, C. *Chromosoma* **98**, 81–85 (1989).
17. Edgar, B. & O'Farrell, P. H. *Cell* **52**, 469–480 (1990).
18. Macdonald, P. M. & Struhl, G. *Nature* **324**, 537–545 (1986).
19. Lehmann, R. thesis, Univ. Tübingen (1985).
20. Strecker, T. R., Kongusuwan, K., Lengyel, J. A. & Merriam, J. R. *Dev Biol.* **113**, 64–74 (1986).

21. Pankratz, M. J., Busch, M., Hoch, M., Seifert, E. & Jäckle, H. *Science* **255**, 986–989 (1992).
22. Rubin, G. M. & Spradling, A. C. *Nucleic Acids Res.* **11**, 6341–6351 (1983).
23. Pirrota, V. *Vectors, A Survey of Molecular Cloning Vectors and their Use* (eds Rodriguez, R. L. & Denhardt, D. T.) 437–456 (Butterworth, Boston, 1988).
24. Theurkauf, W. E., Baum, H., Bo, J. & Wensink, P. C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8477–8481 (1986).
25. Karch, F., Török, I. & Tissieres, A. *J. molec. Biol.* **148**, 219–230 (1981).
26. Spradling, A. C. *Drosophila, A Practical Approach* (ed. Roberts, D. B.) 175–197 (IRL, Oxford, 1986).
27. Struhl, G. *Nature* **338**, 741–744 (1989).
28. Driever, W. & Nüsslein-Volhard, C. *Nature* **337**, 138–143 (1989).
29. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning* (Cold Spring Harbor Laboratory, New York, 1989).
30. Studier, F. W. & Moffat, B. A. *J. molec. Biol.* **189**, 113–130 (1986).
31. Hoch, M., Seifert, E. & Jäckle, H. *EMBO J.* **10**, 2267–2278 (1991).
32. Krasnow, M. A., Saffman, E. E., Kornfeld, K. & Hogness, D. S. *Cell* **57**, 1031–1043 (1989).
33. Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044–1051 (1982).
34. Merrill, P., Sweeton, D. & Wieschaus, E. *Development* **104**, 495–509 (1988).

ACKNOWLEDGEMENTS. We thank H. Schwarz, S. Elend, R. Lehmann, U. Nauber, A. E. Oro, W. Driever, M. Krasnow, M. J. Pankratz, M. Hoch, N. Gerwin, C. Hake, E. Vetter and A. Hassel-Graf for their contribution. This work was supported by the Max Planck Society and the Deutsche Forschungsgemeinschaft. M.R. is a fellow of the Fonds der Chemischen Industrie.

Structure of the pancreatic lipase–procolipase complex

Herman van Tilbeurgh*, Louis Sarda†, Robert Verger‡ & Christian Cambillau*

* LCCMB-CNRS, Faculté de Médecine Nord,
13326 Marseille Cedex 15, France

† Institut de Chimie Biologique, Faculté St-Charles,
13331 Marseille 3, France

‡ Centre de Biochimie et de Biologie Moléculaire, CNRS,
31 chemin Joseph Aiguier, 13402 Marseille, France

INTERFACIAL adsorption of pancreatic lipase is strongly dependent on the physical chemical properties of the lipid surface. These properties are affected by amphiphiles such as phospholipids and bile salts. In the presence of such amphiphiles, lipase binding to the interface requires a protein cofactor, colipase^{1–3}. We obtained crystals of the pancreatic lipase–procolipase complex and solved the structure at 3.04 Å resolution. Here we describe the structure of procolipase, which essentially consists of three ‘fingers’ and is topologically comparable to snake toxins^{4,5}. The tips of the fingers contain most of the hydrophobic amino acids and presumably form the interfacial binding site. Lipase binding occurs at the opposite side to this site and involves polar interactions. Determination of the three-dimensional structure of pancreatic lipase has revealed the presence of two domains: an amino-terminal domain, at residues 1–336 containing the active site and a carboxy-terminal domain at residues 337–449 (ref. 6). Procolipase binds exclusively to the C-terminal domain of lipase. No conformational change in the lipase molecule is induced by the binding of procolipase.

Colipase from one mammalian species activates pancreatic lipases from other species⁷. We obtained crystals of the porcine procolipase–human pancreatic lipase hybrid complex (see Fig. 1 legend) and solved the structure to 3.04 Å resolution using a combination of molecular and isomorphous replacement (Table 1). The procolipase structure could be fitted into the electron density map (Fig. 1a), except for the five N-terminal and two C-terminal amino acids for which there is no electron density. The N-terminal pentapeptide is not necessary for *in vitro* activity and is released during an activation process^{2,3}.

Procolipase is a flattened molecule with dimensions of about 33 × 24 × 16 Å³ (Fig. 2a). The procolipase structure consists mainly of three finger-shaped regions (formed by residues 26–39, 47–64 and 67–87) held together by disulphide bridges. All of the amino-acid chains forming the fingers contain at least one disulphide link to another chain. Of the observed disulphide bridges in the crystal, only two positions agreed with the previous assignment (with links between residues 23–39 and 49–69; ref.

8). The remaining disulphide connections are between amino acids 17–28, 27–61 and 63–87. A fourth finger (residues 15–23) following the N-terminal β -strand, is much shorter and detached from the core of the molecule. Although many residues adopt a β -conformation, a regular β -sheet is not formed. Hydrogen-bonds between strands are only observed close to the disulphide bridges. Amino acids distant from the disulphide links and close to the tips of the fingers are no longer stabilized by hydrogen-bonds and have a poorly defined crystal structure. The lack of electron density for Tyr 55 and the weak density for amino acids 71–75 and 33–34 indicate that these loops are very mobile. The

TABLE 1 Crystallographic data

Data set	Native image plate	Mercuric acetate image plate*†
Resolution limit (Å)	3.04	3.23
Number of measurements	48,860	19,555
Number of unique reflections	17,720	11,416
Completeness (%)	92	74
R_{sym} (%)‡	8.5	7.4

Diffraction data were measured on Image plates (J. Hendrickx and A. Lentfer, EMBL outstation Hamburg, Germany) mounted on the Wigler synchrotron beamline at LURE (Orsay, France) (wavelength 0.95 Å). Images were integrated with the modified MOSFLM programme (A. Leslie, MRC, Cambridge, UK). Further data reduction used the CCP4 suite of programs (Daresbury, England). Molecular replacement and final structure refinement were done with XPLOR (A. Brünger, Yale). The rotation search was done with the refined HPL coordinates. Patterson correlation-refinement¹⁷ selected the rotation solution. Translation search revealed that the space group was $P3_221$. Rigid body refinement brought the initial R -factor of 44% down to 40.9%. For data between 8.0 and 3.5 Å. Simulated annealing refinement gave a final R -factor of 27.1%. At this stage it was not possible to construct the procolipase in the $2F_{\text{obs}} - F_{\text{calc}}$ difference map. Potential heavy-atom derivatives were thus screened to 4.5 Å resolution on a Siemens area detector mounted on a Rigaku rotating anode (data not shown). Crystals were soaked in mother liquid containing the heavy-metal compound. This yielded one useful derivative which was collected to 3.3 Å on image plates (LURE). Phase combination of molecular and isomorphous replacement phases allowed us to construct a colipase backbone model using the program TURBO-FRODO (ref. 18) implemented on a Silicon Graphics 4D/380 graphics system. This partial model was refined to an R -factor of 23% and the protein sequence could then be traced into the electron density map. The model determined at 3.04 Å resolution contains 5,055 atoms and 35 water molecules, and has an R -factor of 17.7%, calculated for all the reflections.

* R_{so} , 18.5% ($\Sigma |I_{\text{deriv}} - I_{\text{native}}| / \Sigma I_{\text{native}}$).

† Phasing power, 1.65 (r.m.s. F_h/E ; F_h , heavy atom structure factor; E , residual lack of closure).

‡ $R_{\text{sym}} = \Sigma |I - \langle I \rangle| / \Sigma \langle I \rangle$.

FIG. 1 *a*, Electron density map at 3.04 Å resolution after final refinement of the model. A hairpin loop of procolipase (comprising residues 44 to 46) that interacts with lipase is represented. The map was calculated with $(2F_{\text{obs}} - F_{\text{calc}})$ Fourier coefficients and contoured at the 1σ level. *b*, Contact region between colipase (thick lines) and lipase (thin lines). Some of the hydrogen-bonds are represented (broken lines).

METHODS. Crystals of the porcine procolipase-human pancreatic lipase (PCL-HPL) hybrid complex were obtained from a solution containing 0.2 M–0.5 M NaCl, 2% polyethylene glycol 8,000, 0.1 M MES (pH 6.0), 8 mg ml⁻¹ HPL and 2.5 mg ml⁻¹ PCL. Crystals first appeared a few hours after adding 1 µl of β-octyl glucoside to a small amount of solution, and grew over about 48 h to ~0.3 × 0.3 × 0.5 mm. The crystals belong to the trigonal space group *P*3₂21 (*a*=*b*=80 Å, *c*=251 Å). The regions around Tyr 55 and Leu 75 have weak electron densities, but are represented for completeness.

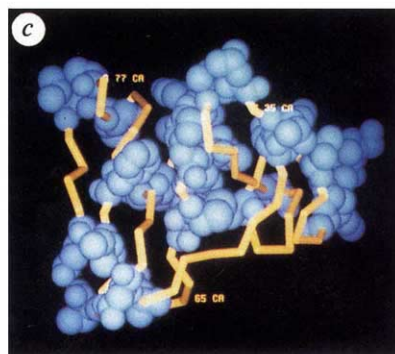
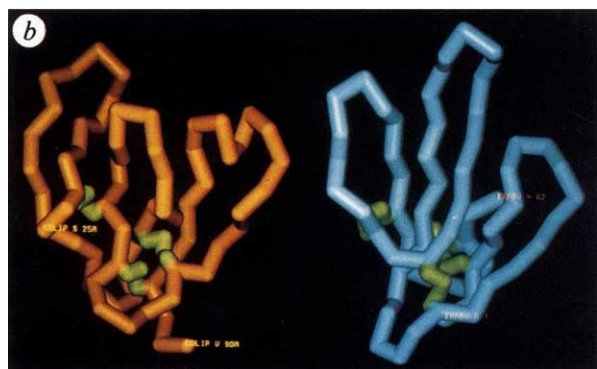
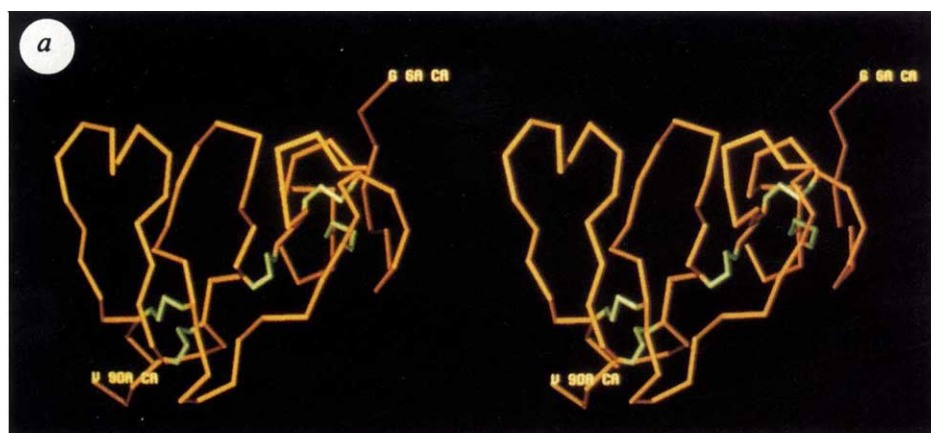
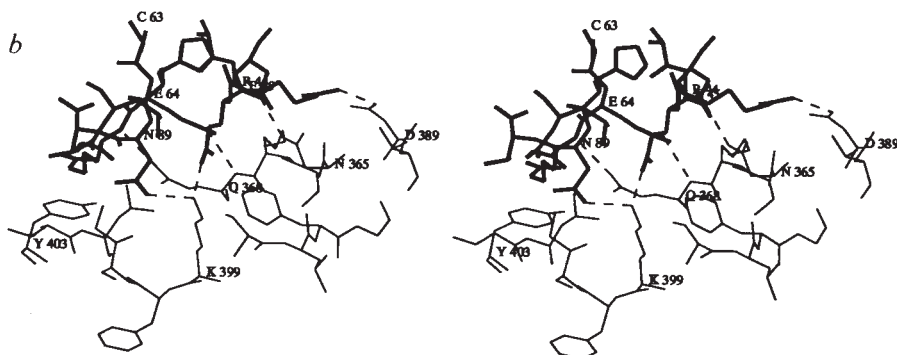
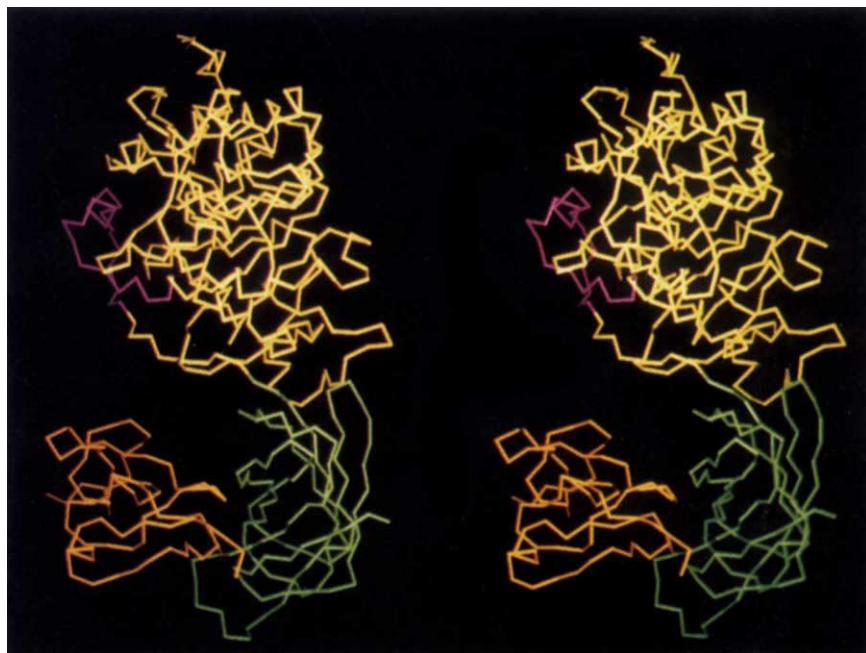


FIG. 2 *a*, Stereo view of the Cα tracing of procolipase. Cysteine bridges are coloured green. The view shows the three hydrophobic fingers in the plane. The lipase-binding site is situated at the bottom. *b*, Structural comparison of the Cα tracings and disulphide bridges (green) in procolipase (orange) and sea-snake venom erabutoxin (ref. 5) (blue). Only part of the procolipase (residues 25–90) is shown. The overall architecture of the two proteins is very similar but the three fingers do not superimpose. *c*, Clusters of hydrophobic amino acids (Ile, Leu, Val, Phe and Tyr) in procolipase are represented as blue spheres. The majority of these amino acids are situated at the tips of the fingers opposite the lipase-binding site. We suggest that this cluster forms the lipid-water interface-binding site.

FIG. 3 Stereo view of the C α tracing of the lipase-procolipase complex. The procolipase molecule is shown in orange, the lipase catalytic domain (residues 1–336) in yellow, the C-terminal β -domain (residues 337–449) in green and the amphipathic flap (residues 242–264), covering the active site, in pink.



overall architecture and the spatial distribution of disulphide bridges bear some resemblance to the structure of erabutoxin from snake venom (Fig. 2b).

Procolipase binds to lipase at its edge with the plane of the procolipase roughly perpendicular to the C-terminal β -sheet domain of the lipase molecule (Fig. 3). Procolipase interacts with the centre of the sheet, which is formed by strands β 2 (residues 352–361), β 3 (364–366 and 369–374) and β 5 (396–403), and Asp 389 (between β 4 (380–388) and β 5) of the C terminus (secondary structure nomenclature according to ref. 6). The lipase binding site is in the regions with the most well defined crystal structure of the procolipase molecule. It is composed of residues from different regions in the procolipase amino-acid sequence; two hairpin loops (residues 44–46, 65–67) and one amino acid (Asn 89) close to the C terminus are interacting with amino acids from various β -strands in the lipase molecule (Fig. 1b). Two salt-bridges (Lys 399–Glu 45, Asp 389–Arg 44; lipase–procolipase, respectively) and six hydrogen-bonds provide the polar component of the binding energy. Apolar binding is mainly conferred by the stacking of Tyr 403 of lipase and Arg 65 of procolipase. In spite of numerous van der Waals contacts (90 contacts in <4.0 Å) there are few interactions between lipase and procolipase. This agrees with the low binding constant ($K_a = 2 \cdot 10^6$ M $^{-1}$) of the procolipase–lipase complex in the absence of an interface⁹. The total buried surface area, calculated with the DDSP program (ref. 10), is 630 Å 2 for the procolipase and 582 Å 2 for lipase. This falls within the range of values reported for various protein–protein complexes¹¹. The lipase–procolipase complex corresponds more to the antibody–antigen-type complexes, where there is complementarity of surfaces, than to the proteinase–inhibitor complexes, where a long loop of the inhibitor protrudes into the active-site groove.

The structure of the complex demonstrates that procolipase binds exclusively to the C-terminal domain of pancreatic lipase. Recent partial proteolysis experiments showed that the intact C-terminal domain, isolated from lipase, was recognized by procolipase¹². In contrast, a proteolytically cleaved lipase in which part of the C terminus was missing was enzymatically active but could not be activated by colipase in the presence of bile salts.

The root mean square difference between the C α -backbone coordinates of the free and procolipase-bound lipase is 0.22 Å. No significant differences, apart from changes in some side-chain conformations, were observed between these two structures. Thus procolipase binds to lipase without inducing a conformational change.

From the crystal structure of the lipase–procolipase complex presented here, a hypothetical lipid-binding site can be located on the procolipase. The majority of hydrophobic amino acids are found in the region opposite the lipase-binding site (Fig. 2c). The peptide segment containing three tyrosine residues (55, 58 and 59), shown by nuclear magnetic resonance and fluorescence spectroscopy to be involved in the interaction with lipid aggregates^{13,14} corresponds to the central finger (residues 47–64) and is flanked by two other fingers. Together they may form an extended hydrophobic surface interacting with the lipid (Fig. 2c). An imaginary plane can be drawn through this hydrophobic region intersecting the surface helix (residues 248–258) in the catalytic domain of lipase (Fig. 3). This helix is part of the amphipathic lid covering the active site catalytic triad. Two structures of *Rhizomucor miehei* lipase, with covalently bound inhibitors, have been recently determined^{15,16}. This microbial lipase also possesses a lid which pivots away from the active site in the inactivated enzyme, to expose a large hydrophobic region to the aqueous solvent and give access to the active site. In our three-dimensional model of the complex, the lipase active site is in its catalytically inactive configuration (lid closed). The binding of pancreatic lipase, but not of colipase, to lipids is impaired by amphiphiles (such as bile salts and phospholipids), which decrease the hydrophobic/hydrophilic balance of the triglyceride–water interface. By forming a specific complex with pancreatic lipase, the protein cofactor brings the enzyme into close contact with the interface and thus indirectly allows the opening of the lid (interfacial activation). □

Received 22 April; accepted 15 July 1992.

1. Verger, R. *Lipases* 83–150 (eds Borgström, B. & Brockman, H.) (Elsevier, Amsterdam, 1984).
2. Borgström, B. & Erlanson-Albertsson, C. in *Lipases* (eds Borgström, B. & Brockman, H.) 151–183 (Elsevier, Amsterdam, 1984).
3. Sarda, L. in *Lipases: Structure, Mechanism and Genetic Engineering* Vol. 16 (eds Alberghina et al.) 145–155 (GBF monographs, VCH, Weinheim, 1991).

4. Rees, B. *et al. Proc. natn Acad. Sci. U.S.A.* **84**, 3132–3136 (1987).
5. Bourne, P. E. *et al. Eur. J. Biochem.* **153**, 521–527 (1985).
6. Winkler, F. K., D'Arcy, A. & Hunziker, W. *Nature* **343**, 771–774 (1990).
7. Rathelot, J. *et al. Biochimie* **63**, 227–234 (1981).
8. Erlanson, C., Charles, M., Astier, M. & Desnuelle, P. *Biochim. biophys. Acta* **259**, 198–203 (1974).
9. Donner, J., Spink, C., Borgström, B. & Sjöholm, I. *Biochemistry* **15**, 5413–5417 (1976).
10. Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).
11. Janin, J. & Chothia, C. *J. biol. Chem.* **266**, 16027–16030 (1990).
12. Abousalham, A., Chaillan, C., Kerfelec, B., Foglizzo, E. & Chapus, C. *Protein Engng.* **5**, 105–111 (1992).
13. Cozzzone, P. J., Canioni, P., Sarda, L. & Kaptein, R. *Eur. J. Biochem.* **114**, 119–126 (1984).
14. McIntyre, J. C., Hundley, P. & Behnke, W. D. *Biochem. J.* **245**, 821–829 (1987).
15. Brzozowski, A. M. *et al. Nature* **351**, 491–494 (1991).
16. Derewenda, U., Brzozowski, A. M., Lawson, D. M. & Derewenda, Z. S. *Biochemistry* **31**, 1532–1541 (1992).
17. Brünger, A. T. *Acta crystallogr.* **46**, 46–57 (1990).
18. Roussel, A. & Cambillau, C. in *Silicon Graphics Directory* (Silicon Graphics, Mountain View, California, 1991).

ACKNOWLEDGEMENTS. We thank the technical staff of LURE, Orsay, for help during data collection, F. Winkler (Basel, Switzerland) for providing coordinates of HPL, M. P. Nésé, N. Ryani, C. Dezan, B. Bellon (Marseille) and Y. Gargouri (Sfax, Tunisia). The study was supported by the EEC BRIDGE-Lipase project, the CNRS, IMABIO programme and the PACA region.

Gene fusion with an *ETS* DNA-binding domain caused by chromosome translocation in human tumours

Olivier Delattre, Jessica Zucman, Béatrice Plougastel, Chantal Desmaze, Thomas Melot, Martine Peter, Heinrich Kovar*, Isabelle Joubert, Pieter de Jong†, Guy Rouleau‡, Alain Aurias & Gilles Thomas§

Laboratoires de Génétique des Tumeurs and URA 620 CNRS,

Institut Curie, 26 rue d'Ulm, 75231 Paris Cedex 05, France

* Children's Cancer Research Institute, St Anna Kinderspital A-1090 Vienna, Austria

† Human Genome Center, Lawrence Livermore National Laboratory, University of California, PO Box 5507, Livermore, California 94550, USA

‡ Centre de Recherche en Neurosciences, McGill University, 1650 Cedar Avenue, Montreal, Quebec H3G 1A4, Canada

§ To whom correspondence should be addressed

EWING'S sarcoma and related subtypes of primitive neuroectodermal tumours share a recurrent and specific t(11;22)(q24;q12) chromosome translocation^{1–8}, the breakpoints of which have recently been cloned⁹. Phylogenetically conserved restriction fragments in the vicinity of EWSR1 and EWSR2, the genomic regions where the breakpoints of chromosome 22 and chromosome 11 are, respectively, have allowed identification of transcribed sequences from these regions and has indicated that a hybrid transcript might be generated by the translocation⁹. Here we use these fragments to screen human complementary DNA libraries to show that the translocation alters the open reading frame of an expressed gene on chromosome 22 gene by substituting a sequence encoding a putative RNA-binding domain for that of the DNA-binding domain of the human homologue of murine *Fli-1*.

Four overlapping cDNA clones that hybridize with probes 22RR3 and 22RR12 located on either side of EWSR1 (Fig. 1)

were isolated. The largest clone contained a 1,968 base pair (bp) open reading frame. It encodes a protein of 656 amino acids which was termed EWS (Fig. 2), the structural characteristics of which are shown in Fig. 3a. The first 285 residues are termed the amino-terminal domain of the EWS protein (NTD-EWS), and contains a degenerate repeated motif with a frequently occurring Ser-Tyr dipeptide (Fig. 3b). The region is also rich in glutamine, threonine and proline residues. A database search (with Fasta in NBRF release 30 and Swissprot release 19) revealed that the region between amino acids 157 and 262 shares 40% homology with the carboxy-terminal domain of the large subunit of eukaryotic RNA polymerase II proteins (CTD-polII)¹⁰. But owing to the degeneracy of the repeats and the lack of proline at crucial positions, it is unlikely that NTD-EWS adopts the regular overlapping β -turns structure predicted for CTD-polII (ref. 11). The remainder of the EWS protein contains three regions rich in glycine, arginine and proline (Fig. 3a) that share homologies with such glycine-rich proteins as single-stranded nucleic-acid binding proteins. Between the first and second regions there is an 85-amino-acid sequence which shares homologies with the RNA-binding domain of several proteins^{12–15} (Fig. 3c). The putative RNA-binding domain of EWS differs from the consensus structure, having an unusually long second loop and a charged amino acid at position three of the consensus motif termed RNP1 (refs 15, 16). These two characteristics are also found in the putative translation product of *Drosophila* cDNA clone p19, a gene with no known function, which shares greatest homology with the C-terminal half of EWS¹⁷. The homologies of the EWS protein indicate that it possesses two different functional domains which could participate in RNA synthesis or processing. The gene *EWS* on chromosome 22 encodes the EWS protein.

Probe 11RR1, adjacent to EWSR2 on chromosome 11 (Fig. 1) was used to retrieve 11 overlapping cDNA clones. The longest clone, BM025, with a 2,939 bp insert, contained a 1,356 bp open reading frame. Analysis of the deduced amino-acid sequence revealed 97% identity with the mouse *Fli-1* gene¹⁸. The retroviral insertion site activating murine *Fli-1* resides on mouse chromosome 9 (ref. 18). This site is phylogenetically conserved and is homologous to a human chromosome 11 region in the proximity

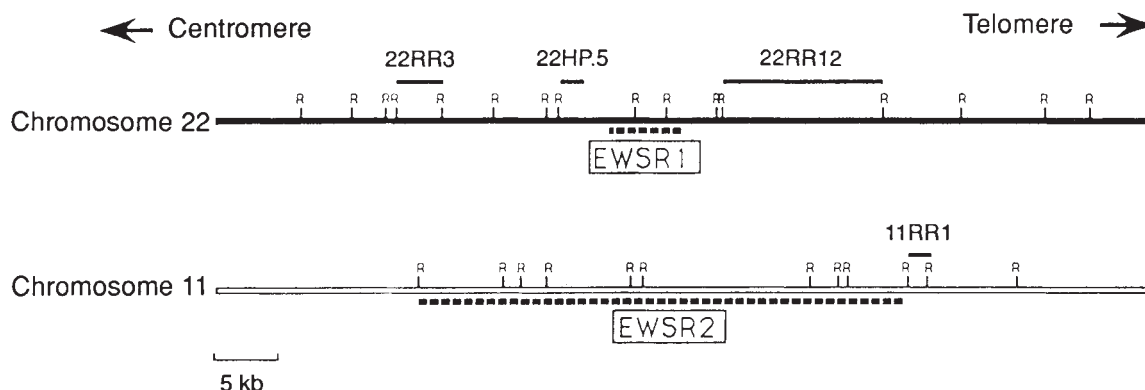


FIG. 1 *EcoRI* restriction map of the EWSR1 and EWSR2 regions. The centromere to telomere orientation and positions of the genomic probes used are indicated.

[illegible]

FIG. 2 Nucleotide and deduced amino-acid sequences (single letter code) of a cDNA containing the entire coding and 3' untranslated regions of *EWS*. The sequences are numbered on the left. Two polyadenylation signals, beginning at nucleotide 2,162 and 2,351 are underlined. The first methionine codon is located with a purine (A) at position -3 and a guanosine (G) at position +4 in accordance with Kozak consensus sequence²⁹ and an in-frame stop codon lies 27 nucleotides upstream. Codons 265 (split in type 1 and type 2 chimaeric transcript) and 349 (split in type 3 chimaeric transcript) are doubly underlined.

METHODS. Probes 22RR3 and 22RR12 (Fig. 1) were used to screen 1×10^6 clones from a human fetal brain cDNA library (Stratagene catalogue number 936206) and identified four clones that hybridized with both probes including clone BF1AC5. Probe 11RR1 (Fig. 1) was used to screen 1×10^6 clones from a human bone marrow cDNA library (Clontech catalogue number

HL1058) and identified 11 clones including BM025. DNA fragments from BF1AC5 and BM025 were subcloned into single-stranded M13mp18 or M13mp19 phages and were used as templates for determination of nucleotide sequences using the dideoxy chain termination method and either a modified T7 polymerase or a *taq* polymerase (Amersham). Complementary sequence information on the 5' and 3' untranslated regions of both genes were obtained from overlapping cDNA clones isolated from cDNA libraries generated from the human breast cell line HBL100 (C. Le Chalony and G. Goubin, unpublished data), from a Ewing's sarcoma cell line (H. Kovar, unpublished data) and from fetal brain (Stratagene catalogue number 937201). The 2,390 bp *EWS* cDNA sequence was determined on both strands of overlapping subclones, using either the M13 primers or custom-made primers.

of the human *ETS-1* gene¹⁹. The homology and syntenic conservation led us to propose that BM025 represents the cDNA of the human homologue of murine *Fli-1* (*Hum-FLI-1*).

Hybridization of the 5' ends of both cDNAs on the cosmid contiguous units of chromosome 11 and 22 (ref. 9) revealed that both genes were transcribed in a centromeric to telomeric orientation. The translocation was shown to be reciprocal on each of the two derivative chromosomes, with no marked loss of genetic material, hence the 5' end of one gene is juxtaposed to the 3' end of the other gene. The putative chimaeric gene generated on the der(11) chromosome is not expressed to a measurable level on northern blots and is not expected to be involved in the tumour phenotype because the der(11) chromosome can secondarily be lost in Ewing's sarcoma^{20,21}. In contrast, the fusion transcript initiated on chromosome 22 and ending on chromosome 11 is readily seen on northern blots².

To investigate the junction of the two genes, we searched for exons flanking EWSR1 proximally and EWSR2 distally. The genomic fragments 22HP.5 and 11RR1 (Fig. 1) that hybridized to *EWS* and *Hum-FLI-1* transcripts, respectively, were sequenced and in each side revealed the presence of one exon. RNAs from various sources were reverse transcribed and amplified by polymerase chain reaction (PCR) using oligonucleotides homologous to these exons. In contrast to RNAs from breast tissue, neuroblastoma (IMR32), pheochromocytoma, lymphoma and ovarian carcinoma, all RNAs from seven primitive

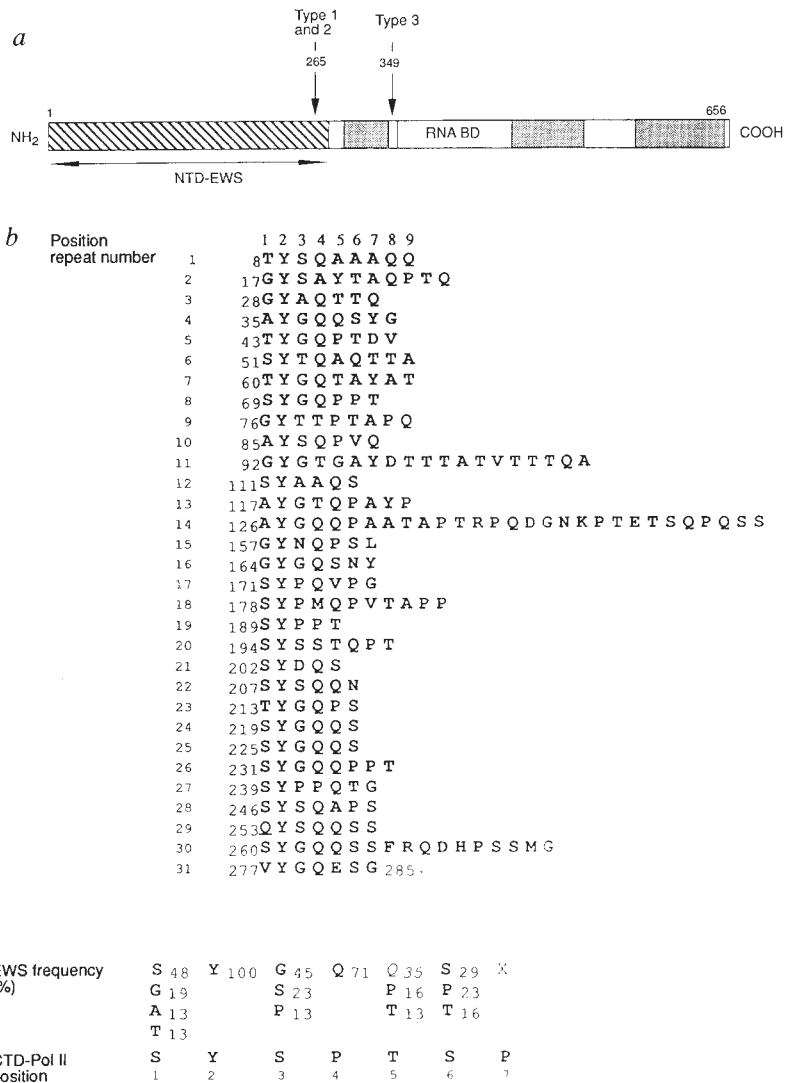
neuroectodermal tumours with t(11; 22) and three unkyarytyped Ewing's Sarcoma tumours promoted the amplification of a specific product. Three different sizes of amplification product were observed, depending on the tumour from which RNA was obtained. (Fig. 4a).

The sequences revealed three types of fusion transcripts. Type one (found five times) contained exonic sequences present in fragments 22HP.5 and 11RR1, and a 174-bp sequence from the adjacent more proximal exon of *Hum-FLI-1* (Fig. 4b). Type two (found four times) and three (found once) differed from type one by the inclusion at the junction site of additional sequences from the coding region of *EWS* and *Hum-FLI-1*, respectively. In all cases the fusion was in-frame and the resulting putative chimaeric proteins differed from the EWS protein by the substitution of the putative RNA-binding domain by the *ETS*-homologous DNA-binding domain of *Hum-FLI-1* (Fig. 4c).

Finally, we isolated a clone containing the entire coding region of the chimaeric messenger RNA from a Ewing's sarcoma cDNA library. Sequencing of this clone showed no additional alterations in the *EWS* and *Hum-FLI-1* moieties.

The translocation described here results in the fusion of two genes that belong to families not previously implicated in human carcinogenesis: the ETS family of DNA-binding proteins and the family of RNA-binding proteins. It further demonstrates that the involvement of the ETS oncogene family in the

FIG. 3 Structural features of the protein encoded by *EWS*. *a*, Different peptide domains are represented. Cross-hatched region represents the first 285 amino acids (the NTD-EWS region) that contain a high proportion (90% of all residues) of tyrosine, glutamine, serine, threonine, glycine, alanine and proline. The three other shaded areas (residues 300–340, 454–513, 559–640) represent glycine- (46%), arginine- (19%) and proline- (13%) rich regions. RNA BD, Putative RNA-binding domain expanded in *c*. Arrows, Position of the junction point with *Hum-FLI-1* for the three different types of chimaeric proteins as deduced from the reverse transcriptase-PCR experiments described in Fig. 4. The positions in the sequence of the two interrupted codons are indicated. *b*, The N-terminal sequence of *EWS* organized to make apparent the presence of a degenerate motif. In this region, tyrosine occurs every 5 to 9 residues and delimits a degenerate motif that is repeated 31 times. The most frequent residues (with their frequency of occurrence) at a given position of the motif are indicated below and compared with the consensus sequence of the eukaryotic CTD-PolII. *c*, The most significant alignments of the *EWS* protein are shown in the region of the putative RNA-binding domain. Open p19, *Drosophila* pen repeats clone p19 (ref. 17); H nucl, human nucleolin³⁰; HsnRNP1 70K, human small nuclear ribonucleoprotein U1 (*M*₁, 70K)³¹; HhnRNP A1 and B1, human heterogeneous nuclear ribonucleoprotein A1 (ref. 32) and B1 (ref. 33); HPABP, human poly(A) binding protein³⁴. The numbers 2, 3 and 4 refer to different RNA-binding domains in the same protein. Invariant positions among these proteins and conservative substitutions between *EWS* and at least four of the six proteins are indicated in bold characters. The different features of secondary structure are described in ref. 13. RNP-1 and RNP-2 refers to the consensus motifs discussed in ref. 14. Identities, percentage of identical residues between *EWS* and each of the displayed RNA-binding domains. On the C-terminal side, the presence of a glycine-rich region (G rich) and/or a basic domain (Bas.D) is indicated.



C		beta-1	loop-1	alpha-1	loop-2	beta-2	loop-3	beta-3	alpha-2	loop-5	beta-4	Identities	
EWS	362	A IYVQC	L NPSVT	EDDLADFFKQCG	VWKNKRTGQPM	IRIYD	KETGKP	K GDATVSY	ED	PPTAKAAVEWF	QGRDFGCGSK	LKVSLA	RKK..G rich + Bas.D
Dpen p19	7	T LFVSG	M DPST	EQDIETHFGAIG	IAKKKRTMKPK	IRLYAN	KETGAS	K GEATVY	DD	CNAQAQSAIEWF	QGRKFNGNA	IKVSLA	QRQ..G rich + Bas.D 47%
H nucl 4	573	T LFVKG	L SEDDT	ENTLKESF--DG	SVR-----	ARIVD	RETGSS	K GFCFVDF	NS	SEDAKAYMS--	DG-RIDGKN	VTLDWA	KPK..G rich + Bas.D 32%
Hs nRNP1 70K	104	T LFVAR	V NYDCT	ESKLRRFEVYG	PIKRI-----	HKVYS-	KRSCKP	R GYAFIEY	EH	ERDMHSAAYKHA	QGRKIDGRR	VLVDVE	RGR... Bas.D 29%
HPABP 3	192	N VYIKN	F GEDMD	DESLKDLF--G	FALS-----	VKVM-	-ESGKS	K GFGFVSF	ER	REDAGKAVDEM	NGKELAGKQ	IYVGRV	QKK
HhnRNP A1 2	106	K LFVGG	I KEDTE	ENHLRDYFEYGG	PIEV-----	LEIMTD	RSGGKK	R GFAFVF	DD	HDSV KIVIQK	YH-TVNGHN	CEVRKA	LSK..G rich + Bas.D 26%
HhnRNP B1 2	113	K LFVGG	I KEDTE	ENHLRDYFEYGG	PIEV-----	LEIMTD	RSGGKK	R GFAFVF	DD	HDSV KIVIQK	YH-TVNGHN	CEVRKA	LSR..G rich + Bas.D 22%
CONSENSUS		RNP-2						RNP-1					

tumorigenic process is not restricted to retroviral-induced haematological malignancies^{22,23}. The translocation seems to have two direct consequences. First, it places the *Hum-FLI-1* DNA-binding domain under the control of an ectopic promoter, the *EWS* promoter which, as shown by northern blot hybridization, does not share the tissue specificity of the *Hum-FLI-1* promoter and may possibly be strong and ubiquitous⁹. As this DNA-binding domain seems to be preserved in the three deduced chimaeric proteins, this aberrant expression may lead to the deregulation of the *Hum-FLI-1* target genes. Second, the translocation links the *Hum-FLI-1* DNA-binding domain to NTD-EWS. The constant involvement of band 22q12 in Ewing's sarcoma and related tumours¹⁻⁸ and more precisely of the EWSR1 region⁹ indicates that NTD-EWS plays an essential and specific role in the tumorigenic process. Its moderate homology with CTD-PolIII suggests that it may disturb gene expression by mimicking, or interfering with, the normal function of CTD-PolIII within the transcription initiation complex.

Its high glutamine and proline contents, which are reminiscent of those found in the activation domain of transcription factors²⁴ may also indicate that it contributes to an aberrant activation of *Hum-FLI-1* target genes.

Karyotypic analyses demonstrating the presence of a t(11; 22) in small round-cell tumours has been used for several years as a diagnostic criterion for Ewing's sarcoma and primitive neuroectodermal tumours. Reverse transcriptase-PCR now provides a sensitive method to demonstrate specific fusion transcripts in these tumours and in their potential metastatic sites. It also becomes possible to document the frequency and specificity of *EWS* and *Hum-FLI-1* alterations in human malignancies, particularly in those with cytogenetically aberrant 22q12 and/or 11q24 regions^{1-8,25-27}. The hypothesized relationship between *EWS* and the inheritable genetic defect causing neurofibromatosis type II, that has been localized in the same 22q12 region²⁸, may also be examined.

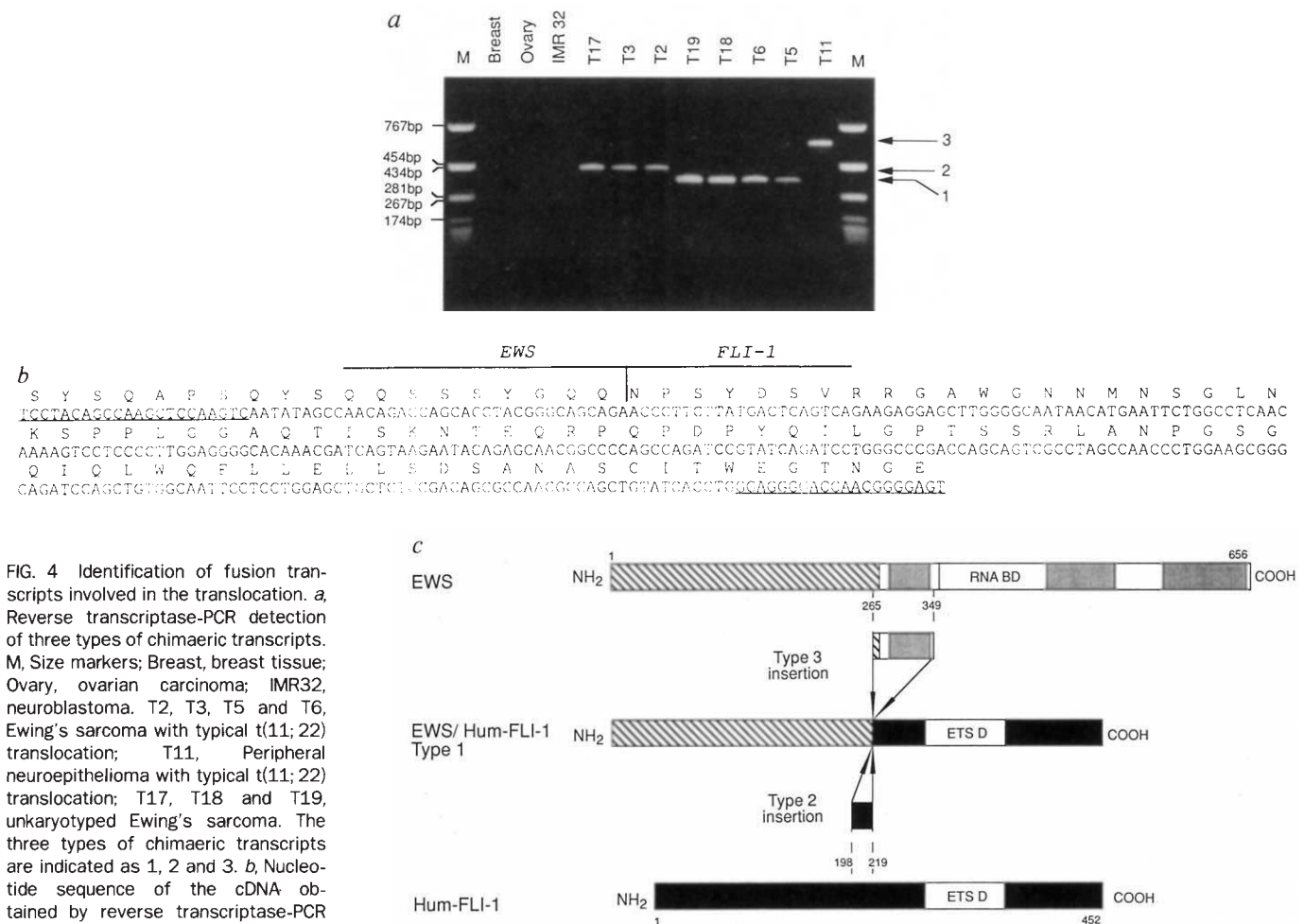


FIG. 4 Identification of fusion transcripts involved in the translocation. **a**, Reverse transcriptase-PCR detection of three types of chimaeric transcripts. M, Size markers; Breast, breast tissue; Ovary, ovarian carcinoma; IMR32, neuroblastoma. T2, T3, T5 and T6, Ewing's sarcoma with typical t(11;22) translocation; T11, Peripheral neuroepithelioma with typical t(11;22) translocation; T17, T18 and T19, untypified Ewing's sarcoma. The three types of chimaeric transcripts are indicated as 1, 2 and 3. **b**, Nucleotide sequence of the cDNA obtained by reverse transcriptase-PCR amplification of the type-1 fusion transcript. Sequences homologous (at the 5' end) or complementary (at the 3' end) to the primers used for the amplification are underlined. The vertical line indicates the junction between the two genes that occurs between the first and second position of codon 265 of *EWS* and between the same positions of codon 219 of *Hum-FLI-1*. **c**, Schematic representations of the proteins encoded by *EWS*, normal *Hum-FLI-1* and their fusion genes. In the three types of chimaeric proteins, the C-terminal portion of *EWS* containing the putative RNA-binding domain (RNA BD) is replaced by the C-terminal part of *Hum-FLI-1* containing the ETS DNA-binding domain (ETS D). The type-1 chimaeric protein is represented entirely schematically. Types 2 or 3 can be deduced from type 1 by the insertion of 84 or 22 additional amino acids from *EWS* or *Hum-FLI-1*, respectively. Positions of

the interrupted codons are indicated in each case.

METHODS. Oligonucleotide 11A (5'-AGAAGGGTACTTGATGG-3') was used as primer for the reverse transcription of 1 mg of total RNA using Gen Amp RNA PCR kit (Cetus). The resulting cDNA was subjected to 30 cycles of PCR amplification with primers 11.3 (5'-ACTCCCGTTGGTCCCTCC-3') and 22.3 (5'-TCCTACAGCCAAGTCCAAGTC-3'), each cycle consisting of a denaturing step at 94 °C for 30 s, annealing at 65 °C for 1 min, and extension at 72 °C for 2 min. The amplified fragment was identified by gel electrophoresis and ethidium bromide staining. Direct sequencing of PCR products was done using Sequenase (USB) after 30 cycles of asymmetrical amplification with either primer 11.3 or 22.3 and after purification through a Centricon 100 membrane (Amicon).

Received 20 May; accepted 28 July 1992.

- Aurias, A., Rimbaut, C., Buffe, C., Dubouset, J. & Mazabraud, A. *N. Engl. J. Med.* **309**, 496-497 (1983).
- Turc-Carel, C., Philip, I., Berge, R. M. P., Philip, T. & Lenoir, G. M. *N. Engl. J. Med.* **309**, 497-498 (1983).
- Wang-Peng, J. *et al. N. Engl. J. Med.* **311**, 584-585 (1984).
- De Chadarevian, J. P., Vekemans, M. & Seemayer, T. A. N. *Engl. J. Med.* **311**, 1702-1703 (1984).
- Vigfusson, N. V., Allen, L. J., Phillips, J. H., Alschibaja, T. & Riches, W. G. *Cancer Genet. Cytogenet.* **22**, 211-218 (1986).
- Cavazzana, A. O. *et al. Adv. Neuroblastoma Res.* **2**, 463-473 (1988).
- Wang-Peng, J., Freter, C. E., Knutsen, T., Nanfro, J. J. & Gazdar, A. *Cancer Genet. Cytogenet.* **29**, 155-157 (1987).
- Turc-Carel, C. *et al. Cancer Genet. Cytogenet.* **32**, 229-238 (1988).
- Zucman, J. *et al. Genes Chrom. Cancer* (in press).
- Corden, J. L. *Trends biochem. Sci.* **15**, 383-387 (1990).
- Suzuki, M. *Nature* **344**, 562-565 (1991).
- Frankel, A. D., Mattaj, I. W. & Rio, D. C. *Cell* **67**, 1041-1046 (1991).
- Kumar, A. *et al. J. Biol. Chem.* **265**, 17094-17100 (1990).
- Munroe, S. H. & Dong, X. *Proc. natn. Acad. Sci. U.S.A.* **89**, 895-899 (1992).
- Kenan, D. J., Query, C. C. & Keene, J. D. *Trends biochem. Sci.* **16**, 214-220 (1991).
- Bandiul, R. J., Swanson, M. S. & Dreyfuss, G. *Genes Dev.* **3**, 431-437 (1989).
- Haynes, S. R., Reibert, M. L., Mozer, B. A., Forquignon, F. & Dawid, I. B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1819-1823 (1987).
- Ben-David, Y., Giddens, E. B., Letwin, K. & Bernstein, A. *Genes Dev.* **5**, 908-918 (1991).
- Baud, V., Lipinski, M., Rassart, E., Poliquin, L. & Bergeron, D. *Genomics* **11**, 223-224 (1991).

- Turc-Carel, C., Philip, I., Berger, M. P., Philip, T. & Lenoir, G. M. *Cancer Genet. Cytogenet.* **12**, 1-19 (1984).
- Douglass, E. C., Valentine, M., Green, A. A., Hayes, F. A. & Thompson, E. I. *J. natn. Cancer Inst.* **77**, 1211-1213 (1986).
- Watson, D. K., Ascione, R. & Papas, T. S. *Crit. Rev. Oncogenesis* **1**, 409-436 (1990).
- Ben-David, Y. & Bernstein, A. *Cell* **66**, 831-834 (1991).
- Mitchell, P. M. & Tjian, R. *Science* **245**, 371-378 (1989).
- Turc-Carel, C., Dal Cin, P., Rao, U., Karakousis, C. & Sandberg, A. *Cancer Genet. Cytogenet.* **30**, 145-150 (1988).
- Shen, W. P. V. *et al. Cancer Genet. Cytogenet.* **45**, 207-217 (1990).
- Trent, J. M., Kaneko, Y. & Mitelman, F. *Cytogenet. Cell Genet.* **51**, 533-562 (1989).
- Rouleau, G. A. *et al. Am. J. hum. Genet.* **46**, 323-328 (1990).
- Kozak, M. *Cell* **44**, 283-292 (1986).
- Srivastava, M., McBride, O. W., Fleming, P. J., Pollard, H. B. & Burns, A. L. *J. biol. Chem.* **265**, 14922-14931 (1990).
- Spritz, R. A. *et al. Nucleic Acids Res.* **15**, 10173-10391 (1987).
- Biamonti, G. *et al. J. mol. Biol.* **207**, 491-503 (1988).
- Burd, C. G., Swanson, M. S., Goerlach, M. & Dreyfuss, G. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9788-9792 (1989).
- Grange, T., Martin de sa, C., Oddos, J. & Pictet, R. *Nucleic Acids Res.* **15**, 4771-4787 (1987).

ACKNOWLEDGEMENTS. We thank P. Dessen for his help in the database search and EWS-structure analysis and A. Mazabraud, J. M. Zucker, J. P. Thierry, C. Petit, D. Chretien, C. Le Chalony and G. Goubin for providing materials. This work was supported by the European Community Commissions, the Ministère de la Recherche et de la Technologie, the Association pour la Recherche sur le Cancer, the Fondation pour la Recherche Médicale and the Council for Coordinating Cancer Research.

GUIDE TO AUTHORS

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without interruptions of the text, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, who are chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are typeset from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose, and to mention availability of these data.

Once a manuscript is accepted for publication, contributors will receive galley proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check their proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few subheadings of two or three words.

Letters are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned but suggestions can be made to the commentary editor in the form of a one-page synopsis. Commentaries are normally between one and four pages of *Nature*.

News and Views articles inform non-specialist readers about new scientific advances, sometimes in the form of a conference report.

Most are commissioned but proposals can be made in advance to the News and Views editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words and 5 references.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and three copies are required, each accompanied by artwork. If photographs are included, four sets of originals are required; for line drawings, one set of originals and three good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Relevant manuscripts in press or submitted for publication elsewhere should be included with each copy of a submitted manuscript, and clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their manuscript numbers.

Titles say what the paper is about with the minimum of technical terminology and in fewer than 80 characters in the case of Articles and Letters. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs. Numbering of sequences should be in the left-hand margin only, with a single space between rows.

Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent the publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference, including unrefereed abstracts, should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included; reference to books should include publisher, place and date.

Abbreviations, symbols, units and greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used except for changed addresses.

Acknowledgements are brief; grant and contribution numbers are not allowed.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1137 National Press Building, Washington, DC 20045, USA. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT). □

Capillary array electrophoresis: an approach to high-speed, high-throughput DNA sequencing

Richard A. Mathies and Xiaohua C. Huang

The speed and throughput of DNA sequencing can be increased by performing sequencing separations on arrays of microcapillaries followed by detection with a laser-excited, confocal-fluorescence scanner.

THE development of high-speed, high-throughput DNA sequencing methods is one prerequisite for the success of the Human Genome Project¹. Curiously, the most powerful separation methods for DNA sequencing, slab and capillary gel electrophoresis, have about the same limiting sequencing rate (~1 kilobase (kb) per hour) and nearly complementary advantages and disadvantages. Although multiple lanes can be run simultaneously on slab gels, they can be tedious to load, the separation is slow because of the low-field strengths typically employed, and DNA-migration and gel-heating anomalies can complicate base reading. Capillary electrophoresis (CE) using gel-filled, narrow-bore (10–100- μm internal diameter) capillaries provides rapid, high-field, high-resolution separations without heating artefacts, requires small sample loads, and has a loading procedure that is easily automated^{2–6}. Capillary electrophoresis separations are complete in about an hour but, because only one capillary can be run at a time, the overall throughput is about the same as that of automated slab-gel sequencing. If all of the positive attributes discussed above could be gathered together in a single method, it would be a major advance. This goal requires miniaturizing the electrophoresis format to provide high-field separations and high-sequencing information

density, developing a sensitive detection system that is compatible with this miniaturization and maintaining the ability to easily load and run multiple samples.

Recently, a laser-excited, confocal-fluorescence gel scanner was developed with the sensitivity and spatial resolution (10–100 μm) required to detect bands on miniaturized sequencing gels⁷. This scanner utilizes an epi-illumination format where the laser is focused on the sample by a microscope objective and the fluorescence is gathered by the same objective followed by confocal detection (Fig. 1). One- and two-colour versions of this scanner have been used along with DNA:dye intercalation complexes to detect double-stranded DNA on agarose gels with picogram sensitivity^{8–10}. This scanner has also been used to detect DNA sequencing gels with a limiting sensitivity of ~10 attomole of fluorescently-labelled DNA per band and a spatial resolution of 10 μm ⁷. Thus, from a detection point of view, the reduction of the physical size of DNA sequencing separations by more than 10 fold is feasible. How can DNA sequencing be miniaturized without making it difficult to load and run multiple samples?

Capillary array electrophoresis

Huang *et al.*¹¹ recently reported a new method called capillary array electrophoresis (CAE) where multiple CE separations are performed in parallel using bundles of microcapillaries. Capillary array electrophoresis is advantageous because the capillaries can be separated and independently manipulated at the inlet for ease of rapid, parallel loading. The 100- μm internal diameter, 200- μm outer diameter, gel-filled capillaries are then bundled at the outlet, providing the high density of 100- μm 'gel lanes' desired for efficient scanning. By simply translating the capillaries back and forth using a motor-driven translation stage, the capillaries in the

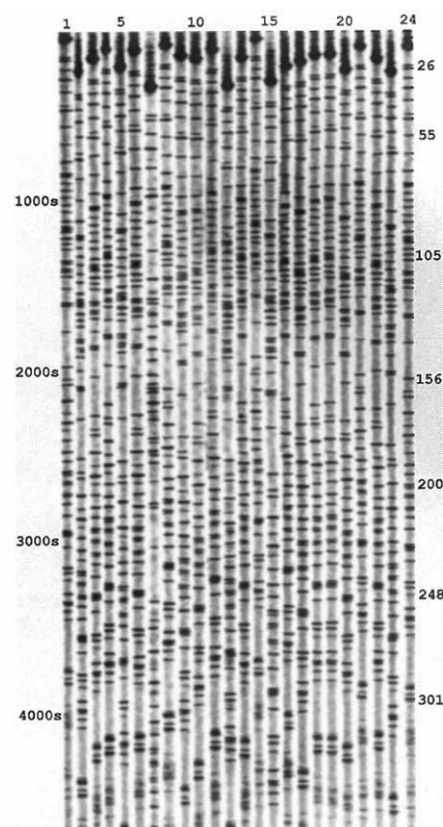


FIG. 2 Image of a separation of M13mp18 T-sequencing fragments performed on a 24-capillary array¹¹.

ribbon are sequentially 'interrogated' by the detection system. As the individual capillaries are only 200- μm wide, hundreds of lanes can be detected by moving the stage only a few centimetres.

Figure 2 presents a separation of fluorescently-labelled M13mp18 T-sequencing fragments performed on an array of 24 capillaries. The vertical dimension of the image gives the elapsed time of the separation while the horizontal dimension shows the 100- μm wide lanes. Note that the entire width of the 24 lanes is less than 5 mm. The separation exhibits excellent resolution and is complete in ~1.5 h at a field strength of ~240 V cm^{-1} . This shows that high-quality separations can be suc-

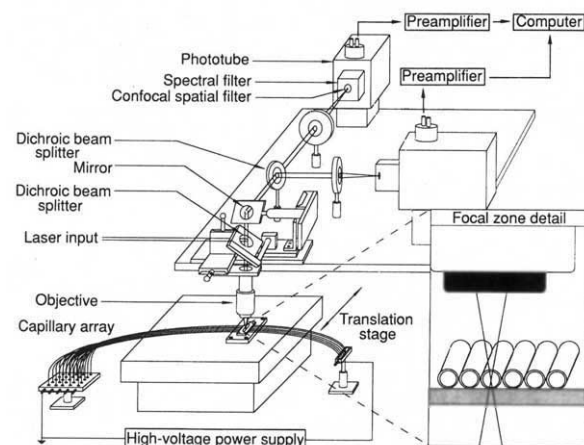


FIG. 1 Capillary array electrophoresis apparatus and two-colour, confocal-fluorescence scanner.

cessfully performed in parallel on capillary arrays and then detected using the confocal scanning system.

The use of CAE for DNA sequencing depends on the ease and reliability with which multiple capillaries can be filled with gel media and run³. Huang *et al.*¹¹ address this issue by using two modifications of the previous CE procedures for DNA sequencing. First, the 9% T, 0% C gels provide noncrosslinked separation matrices that are physically stable and provide excellent separations¹². Second, as a large number of separations are run in parallel, there is no compelling reason to operate the columns at ultra-high fields that can cause breakdown and/or degrade the separation³. Even at a field strength of only $\sim 240 \text{ V cm}^{-1}$, the separation is still ~ 10 -times faster than conventional slab gels.

Binary coding of fragments

For DNA sequencing, the $\sim 5\%$ capillary-to-capillary variation in the migration velocities requires that the four sets of sequencing fragments be separated and detected in the same capillary¹¹. One solution to this problem is to use four different dye-labelled primers (one for each set of sequencing fragments) and a four-colour detection system¹³. This method has been successfully applied to sequencing separations on single capillaries^{4,5}. However, it requires a complex four-colour detection system, as well as correction for the differential migration velocity shift introduced by the four dyes. The multiple-amplitude method¹⁴ is another coding strategy that has been applied to sequencing on single capillaries^{4,15}. However, it can be difficult to determine four amplitudes throughout the run, and this method requires the use of a modified T7 DNA polymerase with Mn^{2+} , which re-

duces the length of the sequencing run¹⁴. Alternatively, we have recently developed a two-dye 'binary-coding' method for labelling the sets of sequencing fragments that allows us to perform DNA sequencing with a two-colour detection system and only two dye-labelled primers¹⁶.

In the binary-coding method, different mole fraction combinations of two dye-labelled primers, having the same priming sequence but tagged with different dyes, are used to differentiate the four sets of DNA fragments. This method is illustrated in Fig. 3, where a '1' indicates that the set of DNA fragments is synthesized with the corresponding dye primer and a '0' denotes the absence of the corresponding labelled dye primer. The dyes used here, JOE and FAM, are fluorescein derivatives obtained from Applied Biosystems (Foster City, California). The {1,1} code indicates that the A-fragments are synthesized with an $\sim 50/50$ mixture of both dye-labelled primers, the {1,0} coding indicates that the G-fragments are synthesized with the JOE-labelled primer and the {0,1} coding indicates that the T-fragments are synthesized with the FAM-labelled primer. C-fragments are not synthesized in the simple binary-coding method although variants of this method have been developed where all four sets of fragments are directly detected (X. C. Huang & R. A. Mathies, manuscript in preparation). Results from one capillary in a CAE sequencing run are presented in Fig. 4. An image of the output from the two detection channels is presented along with one-dimensional electropherograms of the signal integrated across the width of the capillary. Single-base resolution is achieved throughout the run and the ratio of the signal in the red and the green channels provides a reliable indication of the identity of the base fragments.

Prospects

The CAE DNA sequencing experiments performed thus far with a 25-capillary array demonstrate a raw sequencing rate of 6 kb h^{-1} (350 bases per capillary per 1.5 h)¹⁶. The ultimate rate of CAE sequencing is determined by the number of capillaries that can be scanned, the speed of the separation, and the length of the sequencing run. It is currently feasible to run more than 100 capillaries in an array and further devel-

opment of the sequencing chemistry and detection should permit 500 bases to be reliably determined per capillary. Using these estimates, our current apparatus should be capable of sequencing at a rate of $\sim 33 \text{ kb h}^{-1}$.

Two gel-based sequencing methods have been developed that go beyond the 1 kb h^{-1} barrier: capillary array electrophoresis (CAE) and high-field, ultrathin gel electrophoresis (HUGE)¹⁷. In both approaches, the goal is to reduce the physical dimension(s) of the gel to provide high-field separation of multiple samples. Forming the gel is somewhat easier with the HUGE method because the entire gel is cast in one operation. To make CAE convenient, methods must be developed for automated preparation of gel-filled capillary arrays or preformed commercially available gel-filled capillary arrays must be provided. A second important issue for high-throughput DNA sequencing is the 'loading problem'. Here, CAE is advantageous because the individual capillaries can be separated for parallel electrokinetic injection directly from modified microtitre dishes (see Fig. 1). The lane spacing in CAE is now $200 \mu\text{m}$, but further miniaturization should be possible through the use of smaller capillaries, which would increase the overall throughput. Thin slab gels currently use 1-mm lanes¹⁷, which is probably a limit set by the ease of loading as well as 'lane-wander' artefacts. Time will tell as to which method is ultimately most valuable for routine, high-speed, high-throughput DNA sequencing, but the CAE method looks most promising. The recent use of fluorescence-detected CE to analyse DNA mapping fragments and polymerase

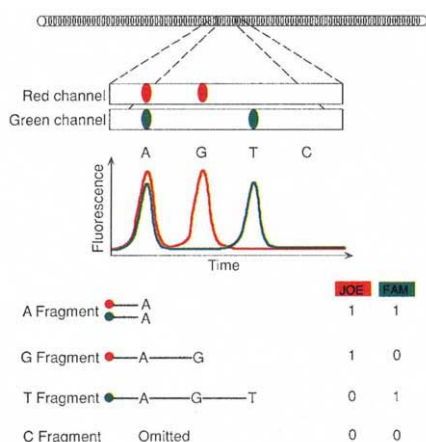


FIG. 3 Binary-coding method for identifying four sets of DNA sequencing fragments in the same capillary by using two dye-labelled primers. The use of only two dye-labelled primers is advantageous because they can be selected to have the same mobility shift effect on the DNA sequencing fragments.

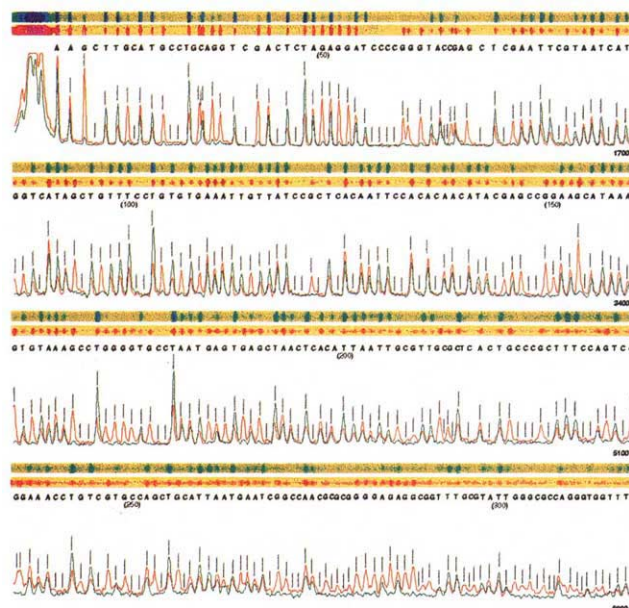


FIG. 4 Analysis of the M13mp18 DNA sequence from one capillary in a capillary array using binary coding. The red image and one-dimensional trace represent the signal detected as a function of time from the red channel. The green image and trace represent the corresponding signal from the green channel¹⁶.

chain reaction products¹⁸ suggests that CAE may also prove valuable in the development of high-speed, high-throughput methods for mapping and diagnostic applications. Miniaturized, massively parallel DNA separation and detection methods are going to be increasingly important in the future development of the Human Genome Project. □

Richard A. Mathies and Xiaohua C. Huang are at the Chemistry Department, University of California, Berkeley, California 94720, USA. This research was supported by the Director, Office of Energy Research, Office of Health and Environmental Research of the US Department of Energy (DOE). Xiaohua Huang was supported by a DOE Human Genome

Distinguished Postdoctoral Fellowship administered by the Oak Ridge Institute for Science and Education. We thank Alexander N. Glazer and Mark A. Quesada for valuable assistance and discussions and John Reilly for preparation of the figures. For more information, fill in reader service number 100.

1. Hunkapiller, T., Kaiser, R. J., Koop, B. F. & Hood, L. *Science* **254**, 59–67 (1991).
2. Karger, B. L. *Nature* **339**, 641–642 (1989).
3. Smith, L. M. *Nature* **349**, 812–813 (1991).
4. Sverdlow, H. et al. *Analyt. Chem.* **63**, 2835–2841 (1991).
5. Luckey, J. A. et al. *Nucleic Acids Res.* **18**, 4417–4421 (1990).
6. Cohen, A. S., Najarian, D. R. & Karger, B. L. *J. Chromatogr.* **516**, 49–60 (1990).
7. Quesada, M. A., Rye, H. S., Gingrich, J. C., Glazer, A. N. & Mathies, R. A. *BioTechniques* **10**,

- 616–625 (1991).
8. Rye, H. S. et al. *Nucleic Acids Res.* **20**, 2803–2812 (1992).
9. Rye, H. S., Quesada, M. A., Peck, K., Mathies, R. A. & Glazer, A. N. *Nucleic Acids Res.* **19**, 327–333 (1991).
10. Glazer, A. N., Peck, K. & Mathies, R. A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3851–3855 (1990).
11. Huang, X. C., Quesada, M. A. & Mathies, R. A. *Analyt. Chem.* **64**, 967–972 (1992).
12. Heiger, D. N., Cohen, A. S. & Karger, B. L. *J. Chromatogr.* **516**, 33–48 (1990).
13. Smith, L. M. et al. *Nature* **321**, 674–679 (1986).
14. Tabor, S. & Richardson, C. C. *J. biol. Chem.* **265**, 8322–8328 (1990).
15. Pentoney, S. L. & Konrad, K. D. *Proc. CLEO/QELS*, 94 (Anaheim, California, May 10–15, 1992).
16. Huang, X. C., Quesada, M. A. & Mathies, R. A. *Analyt. Chem.* **64**, 2149–2154 (1992).
17. Kostichka, A. J., Marchbanks, M. L., Brumley, R. L., Drossman, H. & Smith, L. M. *Bio/Technology* **10**, 78–81 (1992).
18. Schwartz, H. E. & Ulfelder, K. J. *Analyt. Chem.* **64**, 1737–1740 (1992).

Buyer's guide to biotech

This week's profiled products include a biotechnology database, a host of human cDNA clones, a new DNA film scanning system, a protein chemistry station and a preparative two-dimensional electrophoresis system.

PC-ABC — the floppy-disk version of Abstracts in BioCommerce — is a new **biotechnology database for personal computers** from BioCommerce Data (*Reader Service No. 101*). The biotechnology database contains abstracts of over 40,000 articles on biotechnology published in newsletters, newspapers, journals and magazines. Abstracts can be searched by subject using any word(s) in the text, by date or by company name. Full boolean logic is available (for example, 'and', 'or' and 'not' operators) and the software will even find a close match if too specific a selection is keyed in, the company says. PC-ABC is issued twice monthly on a 3.5-inch floppy disk, with quarterly and annual cumulated files included in the subscription. Backfiles from 1981 that reference over 250,000 items can also be purchased. Single-user and Novell-network versions of PC-ABC are available.

Ultra-Violet Products has identified a new application for its Spotcure curing system — the **biotinylation of nucleic acids**

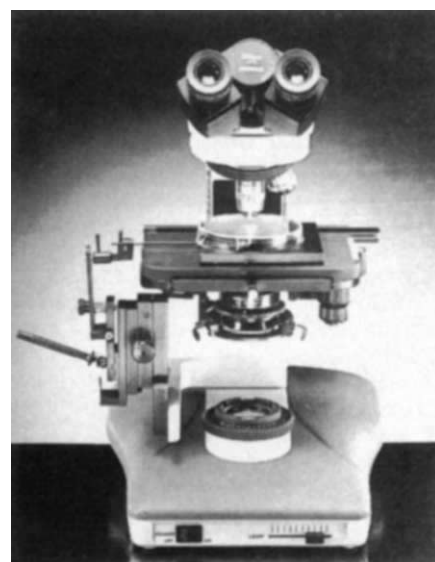


Reaching the spot — see UVP.

(*Reader Service No. 102*). The device allows researchers to direct precisely ultra-violet light at a mixture of nucleic acids and biotin. The light initiates their reaction and generates biotin-labelled molecules. UVP's Spotcure system consists of a lightweight, portable unit that contains a UV lamp and a series of flexible, light-conducting tubes or 'wands' of varying sizes. Two models are available: one that can support up to six wands, the other, two. During the synthesis of an oligonucleotide, for example, UVP says that the primer can be labelled by irradiating the nucleic acid and biotin mixture with UV light. Both DNA and RNA has been labelled successfully, UVP says. Other design features of the Spotcure system include a selection of interchangeable lamps each producing a different wavelength of UV light, manual or automatic control of exposure times and the elimination of unwanted wavelengths by an integral filter.

BioAffinity Systems now has available nitrilotriacetic **agarose** (NTA-adsorbent) as first described in the literature by Hochuli et al. (*J. Chromatogr.* **411**, 177–184; 1987) (*Reader Service No. 103*). The new metal-chelate adsorbent nitrilotriacetic agarose selectively binds protein and peptide targets having adjacent or proximate histidine residues.

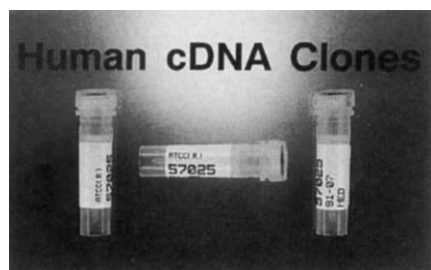
Micro Video Instruments has introduced the Tetrads **dissection microscope**, which is designed specifically for use in yeast genetics (*Reader Service No. 104*). The set-up, which incorporates a Nikon Labophot microscope modified with a fixed-stage, features a joystick micromanipulator and a specimen holder for an inverted Petri dish.



Dissection microscope for yeast genetics.

Optics include planachromat $\times 2$ and $\times 4$ objectives, $\times 15$ wide-field eyepieces and a long-working-distance condenser. A 6-V, 30-W halogen illuminator is included as standard. Additional optics and accessories are available. Features such as the 5-mm detents installed on the y-axis of the mechanical stage are designed to enhance speed and accuracy.

Over 300 **human cDNA clones** from the laboratory of J. Craig Venter formerly of the National Institute of Neurological Disorders and Stroke at the National Institutes of Health are now available through the American Type Culture Collection (*Reader Service No. 105*). The clones are part of Ven-



Assorted human cDNA clones.

ter's effort to identify the approximated 30,000 genes expressed in the human brain. A list of the clones was published in *Science* (252, 1651-1656; 1991). The clone information is also available on the \$25 (US) ATCC Diskette Catalogue of Clones, Vectors, Libraries and Hosts (IBM-compatible format). Sequence information is available from GenBank, T-10, MS K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. Most of the materials are sent as slants of uncharacterized plasmids and approximately 10 per cent are available as characterized clones. The price for each item is \$52 (US) to US and Canadian non-profit organizations; \$80 (US) to other organizations. There is also an additional \$35 (US) preparation fee for slants.

What a transformation!

The Belgian company, Equibio, introduces the new Easyject **twin-pulse electro-injection system**, which can be used with all kinds of cells, from Gram-positive bacteria to fragile eukaryotes (*Reader Service No. 106*). Easyject offers a wide range of adjustable values: resistors (20 Ω to ∞), capacitors (0.5 μF to 3,000 μF) and electric fields (50 to 25,000 V cm^{-1}). Pulse times can be set from between 10 μs and 7 s and experimental parameters can be stored on 'SmartCards'. In addition, the electroporation system offers predetection of arcs, of open and short circuits, and of impedance before, during and after the pulse.

Flowgen is offering the Cellject **electroporation system** for all cell types, including mammalian, plant, yeast and bacterial cells (*Reader Service No. 107*). The Cellject system utilizes a patented double-pulsed system that is designed to reduce cellular shock while increasing transformation efficiencies. All programmed and measured parameters are constantly monitored. These



Cellject electroporation system.

parameters can be output via a serial printer port for hard-copy records. With safety in mind, Cellject detects arc, open and shut circuits, while letting the operator know pre-pulse impedance measurements. Flowgen offers a choice of cuvettes with improved conductance design to decrease further cell mortality.

Image makers

The FilmMate DNA film scanning system from IntelliGenetics combines image analysis software with the automated loading of up to ten films (*Reader Service No. 108*). It is intended to provide an integrated solution to DNA sequence film analysis on the Macintosh. Using 12-bit data processing, FilmMate can distinguish 4,096 shades of grey on a film. With this level of sensitivity, IntelliGenetics says that it can resolve closely spaced bands of variable intensity. Automatic base calling and ambiguity marking with the new BaseWorks software are designed to provide accurate control of sequence identification. The system also

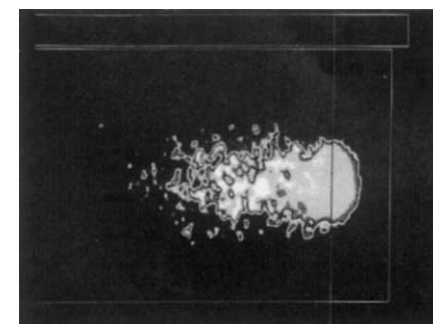
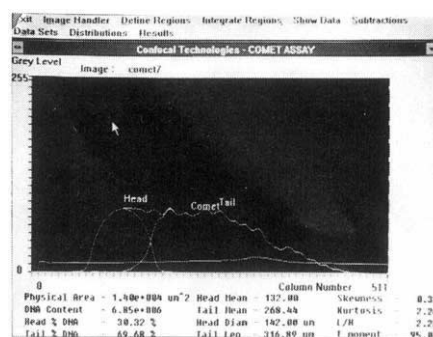


FilmMate DNA film scanning system.

has sequence comparison and alignment capabilities in addition to report formatting options. It can be used with GeneWorks sequence analysis software and Auto-Trans direct transfer electrophoresis systems to provide a complete solution for processing and analysing DNA sequencing samples.

For all your photodocumentation needs, Camag recommends its Reprostar II **transilluminator system**, which can be used with thin-layer chromatograms, electropherograms and gels (*Reader Service No. 109*). The modular photodocumentation system can be combined with various camera formats — Polaroid instant film or conventional film. The new video documentation version features a high-resolution video camera, zoom lens, monitor and video printer.

The 'comet assay' is a new technique for measuring alkali labile sites and overt strand breaks in mammalian cells. Confocal Technologies has been developing automatic image analysis systems to provide rapid and accurate digital solutions to the task of



Analysis screen where 'head' and 'tail' regions of comet are automatically identified and measured (top). Cell in pseudo-coloured micrograph (bottom).

quantifying the cell 'comet' viewed in immunolabelled, epifluorescence microscopy. The company has now released a new version of its Fenestra **digital comet assay system**, which is designed to provide automatic analysis of the geometrical and densitometrical properties of comets (*Reader Service No. 110*). With the new version, essential parameters such as cell length, area and intensity are reported directly into spreadsheet data files. It can be used as a routine evaluation system for genotoxicology laboratories.

Kits

The pGEM-T Vector Systems from Promega allow the **direct cloning of PCR products** without digestion, modification of primers or purification of amplified DNA (*Reader Service No. 111*). The pGEM-T Vector is derived from the pGEM-5Zf(+) Vector, which contains T7 and SP6 RNA polymerase promoters. Stated applications for the pGEM-T Vector include the cloning of PCR products, *in vitro* transcription experiments, the generation of unidirectional deletions, the generation of single-stranded DNA and *in vitro* translation experiments.

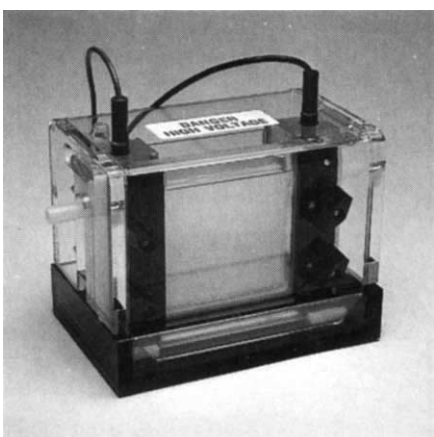
The new **silver staining kit** from E. Merck contains a complete set of reagents for the sensitive detection of proteins and nucleic acids (*Reader Service No. 112*). The advantages of silver staining are, says E. Merck, especially obvious when high-resolution electrophoresis gels are used to which only very small amounts of sample are applied. The sensitivity is 10-100-times higher compared to the use of ethidium

bromide for staining nucleic acids. The company says that reproducible results can be obtained within 2 hours, with sharp band patterns clearly visible on a blank background.

Ideas for instrumentation

Earlier this month, Hewlett-Packard announced the release of a new instrument for the protein chemistry laboratory. The HP G1004A **protein chemistry station** is an automated, benchtop station that concentrates, purifies, fragments and modifies protein before high-performance liquid chromatography or sequence analysis (*Reader Service No. 113*). The sample column allows direct loading of samples from complex matrices or from dilute solutions, which is designed to reduce the number of sample-handling steps and avoid sample loss. Sample solutions may include many contaminants such as salts, nonvolatile buffers, detergents and surfactants, which often interfere with desired chemistries. Chemistries are carried out reproducibly, Hewlett-Packard says, in a closed, inert environment that helps eliminate oxidative side reactions while limiting user exposure to potentially hazardous chemicals. After reaction, sample columns from the protein chemistry station can be moved directly to the HP G1005A sequencer or placed in the HPLC adaptor provided for analysis or fractionation mapping on any HPLC. The protein chemistry station is controlled by the HP 95LX palm-top computer.

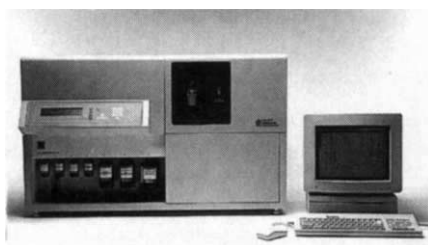
Owl Scientific announces the new Penguin double-sided, **water-cooled electrophoresis apparatus** (*Reader Service No. 114*). The company says that it is the heat distribution and transfer system that allows fast runs with flat, even banding patterns. The central cooling chamber is constructed of thermally-conductive alumina ceramic, which efficiently transfers the heat away from gels. Heat is transferred to cold tap water, which is channelled through the alumina chamber. A third electrode is designed to reduce run times and improve field uniformity. Sliding side clamps apply



Water-cooled gel box.

even pressure on the glass for a leak-proof seal. The platinum electrodes and gold-plated 'banana' jacks are protected against corrosion. The complete set-up includes a Joey Gel Casting System, glass and alumina plates, combs, spacers, clamps and a safety interlocking lid with attached power cords.

The new Model 476 **microsequence analysis system** from Applied Biosystems is designed to meet all your protein sequencing needs in one compact and fully integrated system (*Reader Service No. 115*). The system offers a choice of pulsed-liquid, gas-phase and covalent chemistries. Other design features include 30-min 'FastCycles' and a new Microcartridge, with its reduced volume and surface area, which cuts cycle times by half while enhancing sensitivity,



Microsequence analysis system.

ABI says. This increased sensitivity allows the user to load smaller amounts of sample and take full advantage of such micro-preparative techniques as capillary electrophoresis and microbore HPLC technology. The liquid detection injection system automatically senses the presence of liquid, then triggering injection into the built-in HPLC. Data manipulation, storage and display options can be carried out on the system's Macintosh computer.

Pure and simple

For the rapid purification of genomic DNA from small or large samples of cells, tissue or blood, Invitrogen recommends its TurboGen and Micro-TurboGen **genomic DNA isolation kits** (*Reader Service No. 116*). The protocol utilized by these extractions requires no spooling, incubation with proteolytic enzymes or purification with ionic-exchange resins and yields pure genomic DNA with a molecular weight range of 20–700 kilobases (kb) (average of 300–500 kb). In addition, pipette aspiration is avoided, maintaining the integrity of the DNA, increasing recovery and protecting the researcher from prolonged chloroform exposure, the company says. The TurboGen kit contains all the reagents necessary for 25 isolations of genomic DNA from 10^6 – 10^8 cells, 50–500 mg of tissue or 2 ml of blood in less than an hour. The purified DNA is suitable for polymerase chain reaction (PCR), Southern blotting, restriction fragment length polymorphism mapping and genomic DNA library construction. The Micro-TurboGen kit contains sufficient reagents for 50 isolations of genomic DNA

ADVERTISEMENTS

MICROMANIPULATORS

XENOPUS OOCYTE INJECTION

- Mk 1 "Pantograph" Micromanipulator
- Manual & Automatic Injectors
- Microscopes • Pipette Pullers

SINGER
INSTRUMENTS
"A responsibility to Science"

Call or Fax now for further information

Tel: +44 984 40226 Fax: +44 984 41166

Singer Instrument Co Ltd, Roadwater, Watchet, Somerset, England, TA23 0QL

Reader Service No.12



Severn Biotech Ltd.

CUSTOM SERVICES

- **OLIGONUCLEOTIDE SYNTHESIS**
High quality PROBES, PRIMERS, ANTISENSE OLIGO'S! Fast-Efficient Service offering loads of modifications. BIOTIN, FLUORESCIN, 5' & 3' labelling. Phosphothioates Affinity Support/Oligo's etc.
- **MULTIPLE PEPTIDE SYNTHESIS**
Epitope Mapping & Screening! We now supply 5mg quantities of high purity custom peptides up to 15 residues in length, using Fmoc chemistry. With D or L amino acids, these are supplied lyophilised in 4ml vials, cleaved from solid phase & ready for use.

Tel: 0562 825286 Fax: 0562 825284

Unit 23, Heathcote Enterprise Centre,
Worcester Road, Kidderminster DY9 1DY

Reader Service No.18

Zeineh 1D/2D/DNA/TLC HAND-HELD SCANNING DENSITOMETER

To quantify:

Gels, Blots, TLC, X-Ray film, Photo.

1-D pkg price \$995

8 bits, 60 um resolution, 0-3 OD, sample 11x40 cm. IBM or clones
Options: UV, 2-D, DNA, ELISA, colony counting \$995 each.
Universal software for any scanner. Main 1D, 2D, DNA, \$995 each.
Representatives needed, 30% discount of USA price list.

Do-it-all, Departmental, Laser + Video
Micro Densitometry Station
\$ 16,990



BIOMED INSTRUMENTS INC. 1020 S Raymond Av. #B, Fullerton
CA 92631. Phn: (800) 927-2290, (714) 870-0290, Fax: 17148706385

Reader Service No.22

**Teflon*
Coated
and
Bare Wire from
MEDWIRE**



Medwire provides the highest quality platinum, platinum alloy, silver and stainless steel wire in both single and multistrand configurations.

For prompt delivery or additional information, please contact:

MEDWIRE, an affiliate of Sigmund Cohn Corp.
121 So. Columbus Av. Mount Vernon, NY 10553
Tel: (914) 664-5300 Fax: (914) 664-5377

*Reg. TM of E I DuPont & Co

Reader Service No.25

from 10^4 – 10^5 cells, 50 mg of tissue and 5–100 μ l of whole blood in less than 30 min. The resulting DNA is suitable for PCR and Southern blotting.

The RPM kit that is being distributed by Stratech Scientific is designed for the rapid **isolation and purification of double-stranded DNA** from bacterial cultures (*Reader Service No. 117*). Stratech says that it can be used to deliver high yields of supercoiled plasmid with minimum time and effort. The procedure can be completed in less than 12 min, yielding pure plasmid DNA suitable for enzymatic manipulations, including DNA sequencing, restriction enzyme digestion, *in vitro* transcription and polymerase chain reaction.

By combining the Rotofor preparative isoelectric focusing cell with the Model 491 Prep Cell, Bio-Rad has put together a new **preparative two-dimensional electrophoresis system** that can be used to purify proteins from crude preparations such as cell culture media, tissue homogenates, whole-cell lysates, membrane preparations, sera, plasma, and plant and fungal extracts (*Reader Service No. 118*). Initial fractionation of crude mixtures containing micrograms to 4.0 grams of protein in the Rotofor cell provides up to 500-fold purification of a particular molecule in less than 4 hours, Bio-Rad says. Rotofor



Bio-Rad's preparative 2-D electrophoresis system.

fractions containing proteins of interest can then be applied directly to the Model 491 Prep Cell for a second-dimension purification step to obtain single-band purity of a particular component. Bio-Rad says that proteins can be easily purified for sequencing, antibody production, enzyme kinetics studies and any other *in vitro* and *in vivo* studies where the reagent levels of purified, biologically active protein is required.

[illegible]

Millipore's RFLP analysis software.

from Millipore uses image database technology to assign automatically molecular weights to RFLP and protein band patterns, store images of the band patterns in a database and compare the band patterns in one lane against virtually any number of lanes selected from the database (*Reader Service No. 119*). According to Millipore, matching patterns can be found in minutes. Bio Image RFLP analysis software is available as part of a turnkey Bio Image electrophoresis image analysis system, or it can be added to existing Bio Image systems or Sun Microsystems' computers.

The new 1992-1993 Perkin-Elmer **biotechnology catalogue** from PE Express offers instrumentation, consumables and reagents that are designed to extend the capabilities of the GeneAmp polymerase chain reaction (PCR) process (*Reader Service No. 120*). The catalogue highlights products such as AmpliWax PCR Gem 100, 'multifunctional' recombinant *Thermus thermophilus* DNA polymerase, and thin-walled GeneAmp and Micro-Amp reaction tubes. In addition, it provides information on products developed for new areas of application, including DNA typing and sequencing, and environmental water testing.

The 1992 FMC BioProducts' **catalogue** provides information on a range of agarose products, gel support films and DNA purification products (*Reader Service No. 121*). For added convenience, this year's catalogue includes an agarose selection guide. References, analytical specifications and information on quality and performance tests are also included.

Over 1,000 products are featured in the new 1992-1993 **catalogue** from Clontech (*Reader Service No. 122*). The catalogue includes hundreds of new products, including the human and mouse multiple tissue northern blots (premade northern blots with poly A+ RNA from several human or mouse tissues), the λ DR2 cloning and expression system for efficient direct expression in human cells, polymerase chain reaction (PCR) MIMICs (internal standards for competitive PCR), fluorescein-ON phosphoramidite, and other DNA labelling reagents for automated DNA synthesis, as well as biologically active peptides. $\square$

These notes are compiled by Diane Gershon from information provided by the manufacturers. For more information, fill in the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.

ADVERTISEMENT



Carl Roth GmbH + Co
Schoemperlenstr. 1 – 5
D-7500 Karlsruhe 21
Germany
☎ (07 21) 5 60 60 • Fax (07 21) 56 06 49

Roti®-phenol (equilibrated phenol)
Roti®-phenol / chloroform / isoamylalcohol

*Chemicals + Labware
for Life Science*

